

N70-32267  
Thru  
N70-32283

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

*Space Programs Summary 37-61, Vol. III*

*Supporting Research and Advanced Development*

For the Period December 1, 1969 to January 31, 1970

CASE FILE  
COPY

JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA

February 28, 1970

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

*Space Programs Summary 37-61, Vol. III*

*Supporting Research and Advanced Development*

For the Period December 1, 1969 to January 31, 1970

JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA

February 28, 1970

**SPACE PROGRAMS SUMMARY 37-61, VOL. III**

Copyright © 1970  
Jet Propulsion Laboratory  
California Institute of Technology

Prepared Under Contract No. NAS 7-100  
National Aeronautics and Space Administration

## Preface

The Space Programs Summary is a multivolume, bimonthly publication that presents a review of technical information resulting from current engineering and scientific work performed, or managed, by the Jet Propulsion Laboratory for the National Aeronautics and Space Administration. The Space Programs Summary is currently composed of four volumes.

- Vol. I. *Flight Projects* (Unclassified)
- Vol. II. *The Deep Space Network* (Unclassified)
- Vol. III. *Supporting Research and Advanced Development* (Unclassified)
- Vol. IV. *Flight Projects and Supporting Research and Advanced Development* (Confidential)

# Contents

## ADVANCED STUDIES

<b>I. Future Projects</b> . . . . .	<b>1</b>
A. Geophysical Experiments for the Manned Portion of a Lunar Roving Vehicle Mission	
<i>R. G. Brereton</i> . . . . .	<b>1</b>

## SPACE SCIENCES DIVISION

<b>II. Bioscience</b> . . . . .	<b>5</b>
A. Analysis of the Atmosphere of Jupiter—Requirements for a Combined Atmospheric Gas Chromatographic Mass Spectrometer	
<i>G. P. Shulman and P. G. Simmonds, NASA Code 189-55-03-25</i> . . . . .	<b>5</b>
<b>III. Physics</b> . . . . .	<b>8</b>
A. The Role of CO <sub>2</sub> in the Mars and Venus Atmospheres	
<i>W. B. DeMore and C. Dede, NASA Code 129-01-20-02</i> . . . . .	<b>8</b>
B. Gage Transformation of the Second Post-Newtonian Equations of Hydrodynamics in General Relativity	
<i>F. B. Estabrook, B. K. Harrison, and H. D. Wahlquist, NASA Code 129-04-20-04</i> . . . . .	<b>13</b>
C. An Ion Cyclotron Resonance Study of the Ion-Molecule Reactions in Methane—Ammonia Mixtures	
<i>W. T. Huntress, Jr., and D. D. Elleman, NASA Code 129-02-22-03</i> . . . . .	<b>15</b>
D. Solar Plasma Experiment for Apollo	
<i>C. W. Snyder, D. R. Clay, and M. Neugebauer, NASA Code 914-40-73-07</i> . . . . .	<b>18</b>

## TELECOMMUNICATIONS DIVISION

<b>IV. Communications Systems Research</b> . . . . .	<b>25</b>
A. Coding and Synchronization Studies: Analysis of a Digital Single-Channel Command System	
<i>S. Butman, A. L. Couvillon, R. Goldstein, R. Green, and J. K. Holmes,         NASA Code 125-21-09-03</i> . . . . .	<b>25</b>
B. Coding and Synchronization Studies: Joint Probability Density of True and Estimated Phase Delay in Sequential Ranging System	
<i>S. Butman, NASA Code 125-21-09-03</i> . . . . .	<b>31</b>
C. Coding and Synchronization Studies: On the Behavior of a Phase-Locked Loop During a Temporary Loss of Signal	
<i>U. Timor, NASA Code 125-21-09-03</i> . . . . .	<b>33</b>
D. Coding and Synchronization Studies: The Sensitivity of Sequential Decoders to Noise Correlation	
<i>J. E. Savage, NASA Code 125-21-09-03</i> . . . . .	<b>37</b>

## Contents (contd)

E. Coding and Synchronization Studies: On the Choice of Frequencies for Squarewave Subcarriers C. L. Weber, NASA Code 125-21-09-03 . . . . .	39
F. Coding and Synchronization Studies: Signal Design for Fast-Fading Channels A. V. Balakrishnan, NASA Code 125-21-09-03 . . . . .	42
G. Combinatorial Communications: Error-Correcting Prefixes for Establishing Word Synchronization R. J. McEliece, NASA Code 125-21-09-01 . . . . .	48
H. Combinatorial Communications: Some Results on the Capacity of Graphs R. J. McEliece, R. P. Stanley, and H. Taylor, NASA Code 125-21-09-01 . . . . .	51
I. Combinatorial Communications: Negative Radix Conversion S. Zohar, NASA Code 125-21-09-01 . . . . .	54
J. Propagation Studies: Time-of-Arrival Observations of Eleven Pulsars P. E. Reichley, G. S. Downs, and G. A. Morris, Jr., NASA Code 125-21-09-04 . . . . .	60
K. Propagation Studies: The Solution of a Toeplitz Set of Linear Equations S. Zohar, NASA Code 125-21-09-04 . . . . .	65
L. Multiple Access Communications Research: Correlation With PN Sequences L. D. Baumert and G. A. Morris, Jr., NASA Code 164-21-02-02 . . . . .	73
M. Permutation Transformation of Audio R. M. Goldstein and E. C. Posner . . . . .	77
N. A Class of Sequence Permutations L. R. Welch . . . . .	78
O. Spectrum Analysis of Scrambling Permutation U. Timor . . . . .	81
<b>V. Communications Elements Research . . . . .</b>	<b>90</b>
A. RF Techniques Research: System Studies for Frequencies Above S-Band for Space Communications R. Clauss, NASA Code 125-21-09-08 . . . . .	90
B. RF Techniques Research: S/X-Band Experiment G. Levy, R. Dickinson, and C. Stelzried, NASA Code 125-21-09-08 . . . . .	93
C. Quantum Electronics: Optical Communications Components: 10.6 $\mu\text{m}$ Heterodyne Radiometer C. Finnie, NASA Code 125-21-09-08 . . . . .	95
D. Spacecraft Antenna Research: An RF Study of Reflector Surface Materials for Spacecraft Antennas K. Woo and T. Y. Otoshi, NASA Code 186-68-04-27 . . . . .	99
E. Spacecraft Antenna Research: Radiation Characteristics of Mars Rough Lander Antennas K. Woo, NASA Code 186-68-04-27 . . . . .	106

## Contents (contd)

F. Spacecraft Antenna Research: Derivation of the Geometrical Optics Solution by the WKB Method	
R. Woo, NASA Code 125-21-09-07 . . . . .	109
G. Spacecraft Antenna Research: Spacecraft Antenna Pointing Loss; Optimization Criterion for Illuminating Circular Antenna Apertures	
R. Holland, NASA Code 186-68-04-27 . . . . .	111
<b>VI. Spacecraft Radio . . . . .</b>	<b>121</b>
A. TOPS Radio-Frequency Subsystem	
A. W. Kermode, NASA Code 168-68-04-26 . . . . .	121
<b>VII. Spacecraft Telecommunications Systems . . . . .</b>	<b>126</b>
A. Digital Command System Development	
C. R. Tegnalia, NASA Code 186-68-09-02 . . . . .	126

## GUIDANCE AND CONTROL DIVISION

<b>VIII. Spacecraft Power . . . . .</b>	<b>130</b>
A. A 400-W dc-to-ac Inverter for TOPS Power Subsystem	
D. Hopper, C. McLyman, and J. Loughin, NASA Code 120-60-04-03 . . . . .	130
B. A Test Program for the Investigation of Gravity Effects on Silver-Zinc Cells	
G. L. Juvinall, NASA Code 120-34-10-07 . . . . .	132
C. Charge Recovery and Open-Circuit Voltage Studies of the Liquid Potassium-Amalgam Electrode	
G. L. Juvinall, NASA Code 120-34-10-17 . . . . .	133
D. Solar Cell Research	
P. A. Berman, NASA Code 120-33-14-09 . . . . .	136
E. Evaluation of the Electrical Performance of Mercury-Cadmium Cells	
R. E. Patterson and R. S. Bogner, NASA Code 120-34-10-05 . . . . .	137
F. Electric Propulsion Power Conditioning	
E. N. Costague, NASA Code 120-26-16-04 . . . . .	139
G. Long-Term Space Radiation Environment Testing of Lithium-Diffused Silicon Solar Cells	
J. M. Weingart, NASA Code 120-27-41-06 . . . . .	142
H. Reference Design Criteria for the Thermoelectric Outer-Planet Spacecraft Radioisotope Thermoelectric Generator Power Source	
O. S. Merrill, NASA Code 120-33-14-18 . . . . .	145
I. Documentation Requirements for Obtaining Launch Approval of an Aerospace Nuclear System	
W. B. Weber, NASA Code 120-27-41-06 . . . . .	148

## Contents (contd)

<b>IX. Spacecraft Control</b> . . . . .	149
A. Thermoelectric Outer-Planet Spacecraft Attitude-Control System Simulation Instrumentation <i>L. S. Smith, NASA Code 186-68-02-39</i> . . . . .	149
B. Analysis of Flexible Structure Effects on Attitude-Control Sensors <i>E. L. Marsh</i> . . . . .	151
C. Failure Detection System for Solenoid-Actuated Gas Jet Valves <i>F. G. Roselli-Lorenzini, NASA Code 186-68-02-43</i> . . . . .	155
D. Wiener-Type Filtering for Multivariable Nonlinear Processes <i>A. K. Bejczy, NASA Code 125-17-15-10</i> . . . . .	158

### ENGINEERING MECHANICS DIVISION

<b>X. Materials</b> . . . . .	165
A. Survey of Microstrain Measuring Techniques for Use on Photoheliograph Mirror Material Evaluation <i>W. Rowe, NASA Code 125-22-11-01</i> . . . . .	165
B. Long-Life Spacecraft Pressure Vessel Materials <i>J. C. Lewis, NASA Code 186-68-13-10</i> . . . . .	167
C. Effect of Environment on Spacecraft Thermal Control Materials <i>W. Carroll and J. Crosby, NASA Code 124-09-26-01</i> . . . . .	169
<b>XI. Applied Mechanics</b> . . . . .	171
A. Optimization of Space Antenna Concepts <i>J. N. Yang, NASA Code 124-08-26-01</i> . . . . .	171
B. Receptance Coupling Program <i>R. D. Simpson, NASA Code 124-08-26-02</i> . . . . .	178
C. On the Maximum Dynamic Response of Structures <i>M. Shinozuka, NASA Code 124-08-26-02</i> . . . . .	179
D. Deformation Modes of a Simple Rotating Structure <i>E. O. Weiner, NASA Code 120-13-16-04</i> . . . . .	184
E. Vibration and Deformation Measurements Using Holography Techniques <i>E. Heer, G. Morse, and R. Badin, NASA Code 124-08-26-04</i> . . . . .	187

### PROPULSION DIVISION

<b>XII. Solid Propellant Engineering</b> . . . . .	191
A. Heat Transfer in a Burning Solid Propellant With Oscillating Surface Temperature and Surface Regression <i>R. L. Klaus, NASA Code 128-32-50-01</i> . . . . .	191

## Contents (contd)

B. Experimental Study of the Combustion Termination of a Solid Propellant by Rapid Depressurization W. Gerber, NASA Code 128-32-60-01 . . . . .	200
C. Propulsion System Analysis for Thrusters of Differing Performance Characteristics D. J. Norton, NASA Code 180-32-04-01 . . . . .	207
<b>XIII. Polymer Research . . . . .</b>	<b>213</b>
A. Investigation of Sterilizable Battery Separator Membranes E. F. Cuddihy, D. E. Walmsley, and J. Moacanin, NASA Code 120-34-10-03 . . . . .	213
B. Parallel-Plate Viscometer R. F. Fedors and R. F. Landel, NASA Code 128-32-43-01 . . . . .	216
C. Relationship Between Index of Refraction and Molecular Weight R. A. Rhein and D. D. Lawson, NASA Code 128-32-80-06 . . . . .	219
D. Refractive Index to Molecular Weight Relationships for Isobutylene Telomers and Polymers R. A. Rhein, NASA Code 128-32-80-06 . . . . .	222
E. Electronic Conductivity of Elastomeric Ionomers R. Somoano, S. P. S. Yen, and A. Rembaum, NASA Code 129-03-22-01 . . . . .	226
F. Viscoelastic Behavior of Elastomers Undergoing Scission Reactions J. Moacanin, J. J. Aklonis, and R. F. Landel, NASA Code 129-03-22-02 . . . . .	235
G. Long Term Aging of Elastomers: Chemical Stress Relaxation and Other Studies of Peroxide-Cured Styrene-Butadiene Rubber R. Rakutis, S. H. Kalfayan, and R. H. Silver, NASA Code 129-03-22-02 . . . . .	237
<b>XIV. Research and Advanced Concepts . . . . .</b>	<b>242</b>
A. Comparison of Theoretical and Experimental Deflection of a Blown Arc T. K. Bose, M. B. Noel, and P. F. Massier, NASA Code 129-01-21-02 . . . . .	242
B. Liquid-Metal MHD Power Conversion D. G. Elliott, D. J. Cerini, L. G. Hays, and D. W. Bogdanoff, NASA Code 120-27-42-01 . . . . .	246
<b>XV. Liquid Propulsion . . . . .</b>	<b>250</b>
A. Material Compatibility O. F. Keller and L. R. Toth, NASA Code 731-13-44-02 . . . . .	250

## MISSION ANALYSIS DIVISION

<b>XVI. Systems Analysis Research . . . . .</b>	<b>252</b>
A. Advances in the Numerical Integration of the Motion of the Moon D. B. Holdridge, K. K. Garthwaite, and J. D. Mulholland, NASA Code . . . . .	252
B. Forward Smoothing of Sequentially Correlated Processes T. Nishimura, NASA Code 129-04-20-02 . . . . .	253

## Contents (contd)

C. The Calculation of Probabilities of Mission Failure for Solar-Electric Propulsion Systems Utilizing Varying Numbers of Thrusters <i>W. Kizner, NASA Code 120-26-16-03</i> . . . . .	257
D. Einstein's General Theory of Relativity: Measuring the Sun's Angular Momentum <i>R. W. Davies and H. Lass, NASA Code 129-04-20-02</i> . . . . .	259
<b>Subject Index</b> . . . . .	263

# I. Future Projects

## ADVANCED STUDIES

### A. Geophysical Experiments for the Manned Portion of a Lunar Roving Vehicle Mission,<sup>1</sup>

R. G. Brereton

#### 1. Introduction

The follow-on series of *Apollo* lunar missions, starting with *Apollo 20*, are expected to extend the period of lunar exploration through the mid-1970s. These missions will exploit *Apollo* technology towards the achievement of established lunar program goals. Generally, they will be oriented to provide new and perhaps unique information that will further our understanding of the moon as it is today, its origin and evolution, its structural and age relationship to the earth, and, finally, how it may be utilized as man's new frontier. Roving vehicle technology will be required for these missions to serve as both a mobility aid to the astronaut and as a long-range automated reconnaissance system. Such a dual-mode vehicle is presently under study at JPL and at the Marshall Space Flight Center (MSFC). The lunar roving vehicle criteria

presented in this subsection were prepared for the Joint Panel for Lunar Mobile Science and Traverse Planning (sponsored by NASA through MSFC) as a guide for planning so-called hard-rock geophysical experiments that might be accomplished during the manned portion of the dual-mode roving vehicle mission.

The dual-mode lunar roving vehicle will be a self-contained, electrically-powered vehicle. In the manned mode, the rover will be capable of carrying one full-suited and equipped astronaut plus several hundred pounds of scientific equipment and surface samples. The rover will provide a significant capability for manned lunar exploration since its speed and range will allow time for several sorties into the site environs. The capability to range out from the lunar module in time and distance will not be limited by the operational capability of the rover, but rather by the hours of support that can be provided by the astronaut's portable life-support system.

Voice and science data telemetry will be directed to earth from the rover via an extra-vehicular communications system. This unit will be both battery powered and

---

<sup>1</sup>The author wishes to acknowledge the assistance of H. Ackermann, D. Elston, and R. Dubois in preparing the active seismic and magnetometry portions of this article.

capable of using rover power through a plug-in unit. The portability of the communication unit in the battery-powered mode will extend the range and versatility of the astronaut's surface activities away from the rover. Communications to the communications unit by the astronaut or from scientific experiments will normally be by an RF link, but some hard-line, high-data-rate capability is expected to be available.

The established scientific objectives for the manned portion of the rover mission should be those that make maximum use of the man—and his ability to observe, infer, and do—and a versatile mobility aid. Some of the scientific objectives and tasks for the manned portion of the rover mission are as follows:

- (1) **Geology.** To study basic geologic relations in sites representative of major lunar features and provinces so as to reconstruct the lunar geologic history and processes. The procedures here will be largely those of field geology that can be accomplished by the astronaut and supplemented by observers on earth.
- (2) **Geophysics.** To use geophysical techniques for studying the crustal and subcrustal properties of the moon and to supplement geologic observations. The general class of experiments considered here are gravity, active seismic, magnetism, and electrical.
- (3) **Samples.** To provide representative samples for subsequent analysis on earth, to support the field geology investigation, and to provide the background for later traverse analysis.
- (4) **Bioscience.** To search for any evidence of lunar life forms that are either existent or fossil.
- (5) **Geodesy/cartography.** To refine the figure of the moon and to determine the topographic character of local terrain.
- (6) **Complicated experiments.** To carry out measurements or emplace experiments that cannot be done by automated techniques; for example, astronomy experiments requiring long or oriented antennas and fields and particles experiments requiring special sensor emplacement or long arrays.

## 2. Science Objectives and Requirements

The specific scientific objectives and requirements for gravity, active seismic, magnetometry, and electrical prospecting experiments that may be accomplished as a part of the manned portion of a dual-mode roving vehicle mission to the moon are discussed in the remainder of this article.

### a. Gravimetry experiment.

**Objectives.** The objectives of the gravimetry experiment (SPS 37-57, Vol. III, pp. 1-6) are to provide information on the anomalous gravity field for such lunar features as impact craters, volcanic forms, mascons, ridges, faults, rills, basins, etc. The experiment will also answer such questions as: How does the gravity field for these features compare with terrestrial analogs? What is the density of lunar surface and near-subsurface materials? Can lunar stratigraphic units be mapped by gravity survey techniques? This experiment will also provide information on the lunar crust and isostasy and information towards the task of lunar geodesy; that is, determining the size and shape of the moon, providing control points for mapping and charting work, and providing a gravity base.

**Measurement requirements.** Values of absolute  $g$  to  $\pm 1$  milligal are desirable. (The sensitivity of the gravimeter for a site survey should be  $\pm 0.1$  milligal.) In the event that absolute gravity measurements are not practical, the selected gravimeter should have a calibration error (from earth to moon) of better than  $\pm 3$  milligals. The drift on any instrument used for tidal studies must be less than 0.1 milligal/mo; also, drift during the survey operation should be less than 0.1 milligal. This last factor covers mechanical drift as well as thermal. Gradient measurements, where required, should be made to at least 1 Eötvös unit. Approximately 5 min should be allowed for stabilizing and automatic leveling of the gravimeter after deployment before measurements are made. Measured gravity values should be both displayed at the instrument for easy reading by the astronaut and routed to a telemetry interface for transmission.

**Other requirements.** The system should be portable and self-contained in a configuration capable of being carried by the astronaut. It should have a power source capable of providing power for approximately 4 h of operation, have an automatic stabilizing and leveling system, and weigh less than 30 lb. Also, the system should be capable of being mounted on the rover and automatically lowered to the lunar surface during stops.

The combined free-air and Bouguer effect for the moon is approximately 0.075 milligal/m. In order to keep this effect at a minimum, it is desirable to know elevations over the site to  $\pm 1$  m. For surveys out to the maximum traverse distance, this requirement may be relaxed to  $\pm 10$  m.

The navigation and positioning requirement should include the capability of establishing a position to within

5% of the distance traveled from the lunar module, or any previous reference point. It will also be necessary to know the position of topographic features of more than 10 m in elevation that are at a distance less than 30 m from the gravity sensor to a ranging accuracy of  $\pm 1$  m.

#### **b. Active seismic experiment.**

**Objectives.** The objectives of the active seismic experiment (SPS 37-56, Vol. III, pp. 213-216) are to determine models of subsurface lunar structure from seismic recordings using artificial or natural energy sources. The scale of experiments may vary between shallow soundings for obtaining regolith thicknesses to several tens of kilometers or greater using larger energy sources such as explosives or possible moonquakes.

In addition, the experiment will obtain wave velocities and attenuation constants of lunar subsurface materials and convert these to estimates of their densities, elastic constants and physical properties.

**Measurement requirements.** Measurement requirements for the seismic experiment are as follows:

- (1) **Sensitivity.** A seismic system with the ability to detect ground displacements of greater than  $10^{-7}$  cm in the range of 1 to 30 Hz.
- (2) **Deployment of seismometers.** Seismometer systems of less than 1-kg mass are to be deployed and implanted firmly into the ground. This task can be performed by the astronaut or automatically by the roving vehicle.
- (3) **Seismic energy source.** For shallow soundings, the system should have the ability to impact the moon with roughly 200 ft-lb of energy without imparting energy within 40 dB of this during the descent motion. This impact must be timed to  $\pm 0.001$  s. For deeper soundings, it should be able to deploy a chemical explosive and detonate it with a signal from earth or the rover. The shot instant must be timed to  $\pm 0.004$  s.

**Other requirements.** The system should be portable and self-contained, in a configuration capable of being carried by the astronaut, and weigh less than 50 lb. Specified missions in which deep probing is desirable may require a large energy source that cannot be obtained within this weight limitation.

The navigation and positioning requirement includes the ability to determine relative positions between any loca-

tions occupied by seismometers or energy source to within 5% of the distance between them. This distance may be as small as a few meters or as great as the entire traverse length.

The system must have the data-transmission capability to handle high data rate signals to earth during selected time intervals. Data storage, compression, and subsequent transmission may be required.

#### **c. Traverse magnetometry experiment**

**Objectives.** The objectives of this experiment are to measure values of magnetic vector field, gradient vector field, and magnetic susceptibility in order to determine the size distribution, geometry, location, attitude, and nature of rock masses making up the surface and subsurface of the moon. In addition, this experiment will also measure the areal and temporal variations of the magnetic field at discrete locations on the moon.

**Measurement requirements.** Measurement requirements for the magnetometry experiment are as follows:

- (1) Values of vector field to  $\pm 1$  gamma.
- (2) Values of vector gradient field to  $\pm 0.1$  gamma.
- (3) Magnitude of magnetic susceptibility *in situ* to  $\pm 1 \times 10^{-5}$  emu/cm<sup>3</sup>.
- (4) Gradient field data are to be calibrated in terms of direction and intensity of remanent magnetism (to be used under conditions where applicable)—direction to  $\pm 10$  deg and intensity to  $\pm 1$  order of magnitude.
- (5) Measurements should be displayed at the instrument for easy reading by the astronaut and routed to a telemetry interface for transmission.

**Other requirements.** The system is to be configured to be carried by the astronaut and weigh less than 5 lb. It is to be self-contained and provide a reference orientation (azimuth  $\pm 3$ -deg vertical position sensor) to indicate that the staff is within 5 deg of vertical; the system will also require a sensor for indicating contact with surface. The data transmission capability is to include a power source capable of providing power for 4 h of remote operation. The system should provide meters showing vector sum of gradient field and magnitude of magnetic susceptibility for read-out by the astronaut.

The system is to be capable of being mounted on the rover using an approximately 5-ft-long boom and carried

in a vertical position during traverse. In the mounted configuration, the instrument is to be capable of being lowered to the lunar surface during stops. Measurements are to be made in both the raised and lowered positions.

The navigation and positioning requirement is to include the capability of establishing a position to within 5% of the distance traveled from the lunar module, or any previous reference point, and the capability of determining the position of selected features to the same survey accuracy as established for the field geology surveying requirements.

*d. Electrical surveys.* A variety of experiments and techniques can be considered here. There are electrical prospecting methods that use naturally occurring ground currents and there are others that rely upon artificial currents. Frequencies ranging from direct current to perhaps 1000 Hz have been used. Sometimes the current or potential difference is applied directly to the ground under investigation, while, at other times, electromagnetic waves are generated from a source above the ground and their change, as an effect of the ground, is observed. These experiments are powerful tools in terrestrial prospecting and they have been extensively developed. They will have wide application in the lunar program. Some of the general science objectives and requirements for this class of experiments are described below.

*Objectives.* One of the objectives of this experiment is to determine models of subsurface lunar structure, stratification, and composition using both natural and artificial energy sources. The scale of the experiment may vary depending upon the energy source, specific type of electrical experiment, and the geological and physical properties of lunar rocks. Other objectives are to measure telluric and other natural currents in the ground in order to understand how they are generated in the moon (and to provide a basis for future electrical surveys), investigate the *in situ* electrical properties of lunar material, and to

provide a basis for electrical prospecting for strategic material, e.g., water.

*Measurement requirements.* This experiment may measure the electric potential difference or current in the ground between electrodes. From these data, the conductivity or resistivity, which may range from 1 to  $10^{12}$   $\Omega$ -cm for lunar rock units, and the dielectric constant, which may range from about 10 esu for dry lunar rock to perhaps 50 esu for water-saturated rocks, can be calculated. This experiment may also measure the magnitude and direction of electromagnetic fields induced in lunar material by an external alternating field. The depth penetration of the inducing field is a function of its frequency and the resistivity of the ground. Frequencies in the range of 10 Hz to perhaps 500 Hz in a resistivity medium of about  $10^6$   $\Omega$ -cm are probably the most useful.

The current electrodes or energy source as well as the potential electrodes must be placed in contact with the surface or arranged in an operating position. Astronaut participation in laying out arrays will be required since it will be necessary to have cable contact between the elements of an experiment.

*Other requirements.* The system should be portable and self-contained in a configuration capable of being carried by the astronaut. It should weigh less than 30 lb.

The navigation and positioning requirement is to include the ability to determine relative positions between any locations occupied by electrodes or an energy source to within 5% of the distance between them. This distance may be as small as a few meters or as great as the entire traverse length.

The system must have the data-transmission capability to handle high data rate signals to earth during selected time intervals. Data storage, compression, and subsequent transmission may be required.

## II. Bioscience

### SPACE SCIENCES DIVISION

#### A. Analysis of the Atmosphere of Jupiter— Requirements for a Combined Atmospheric Gas Chromatographic Mass Spectrometer, *G. P. Shulman and P. G. Simmonds*

##### 1. Introduction

Jupiter probes containing analytical instrumentation to determine the composition of the atmosphere are under consideration for missions in the late 1970s. A mass spectrometer with a range of 1–40 amu is the instrument generally specified for this purpose (Ref. 1). Consideration of the science requirements indicates that such a mass spectrometer will be inadequate to yield the data required from such a mission for resolution of the scientific problems. As a minimum, addition of a simple atmospheric gas chromatograph, or preferably, integration of the atmospheric gas chromatograph with the mass spectrometer, would greatly augment the capability of the mission. By 1975, it is anticipated that a total payload weight of about 7 lb would be adequate for such systems.

##### 2. Scientific Rationale

Analysis of the elemental composition of Jupiter will possibly provide information regarding stellar composition (Ref. 1) and/or the composition of the primordial nebula (Ref. 2). Determination of the relative abundance and isotopic composition of all possible elements may thus

shed light on fundamental cosmological questions such as the formation mechanism of the primordial nebula. A knowledge of hydrogen/helium ratios and of their isotopic composition is of primary importance. Similar isotopic composition measurements on the other inert gases (particularly  $A^{36}/A^{40}$ ) have been suggested as a definitive indicator of the history of a planet.<sup>1</sup>

Unfortunately, and contrary to a geocentrically generated impression held by many scientists, a low-resolution mass spectrometer has limited utility for analysis of unknown atmospheres. In part, this is because the relative abundances of the isotopes of a given element are not necessarily identical to the terrestrial values, due to, perhaps, radiological or biological alteration of cosmic abundances. For example, analysis of a hydrogen–helium mixture will give peaks at the following mass-to-charge  $m/e$  ratios:

$m/e$  1 (H)

$m/e$  2 ( $H_2$ , D)

$m/e$  3 (HD,  $He^3$ ,  $H_3$ )

$m/e$  4 ( $D_2$ ,  $He^4$ )

<sup>1</sup>Fanale, F., Stemberge, C. H., and Horowitz, N. H., "Biological Studies of Mars: Theoretical Considerations and a Practical Experiment" (to be published in *AAS Sci. Technol.*).

If the relative abundance of either  $\text{He}^3$  to  $\text{He}^4$  or H to D is not known, the peak at  $m/e$  3 is useless in determining the ratio of  $\text{He}^3$  to HD. Similarly, in a mixture of methane, ammonia, and water, accurate quantitation is impossible unless isotope ratios are known. This is illustrated in Table 1, which shows interference by fragments of each compound in the determination of other compounds. Equally important, if the mass spectrometer is flooded with a major component, minor peaks are not visible. Since flight instrumentation generally has a dynamic range of  $10^5$ , no single isotope present at less than 10 parts/ $10^6$  could be detected. For a typical model (Ref. 1) of the ammonia clouds of Jupiter, this would limit the analysis to hydrogen, helium, methane, neon, ammonia, and the major argon isotope. Thus, it is essential to perform quantitative analyses not only on the major constituents using a mass spectrometer, but also to utilize a gas chromatograph for analysis of components in the fractional part/ $10^6$  range and for separation of a purified sample of each atmospheric constituent prior to its introduction into the mass spectrometer. Under these conditions, isotopic analyses can be performed even on the minor constituents.

Other problems of importance, such as estimation of the extent of convective mixing of the atmosphere, and whether equilibrium or rate-dependent processes control the composition could be attacked by securing analyses as a function of elevation. In particular, analysis for nitrogen would be a sensitive indicator of the relative importance of photolytic dissociation of ammonia to the postulated high-temperature-high-pressure synthesis from nitrogen and hydrogen in the lower atmosphere. Controversy over the presence of organic molecules in the atmosphere is developing (Refs. 3 and 4). While it is not proposed that a specialized organic analysis experiment be developed, extension of the range of the mass spec-

trometer to  $m/e$  60 would allow not only for detection of more complex or higher molecular weight inorganic species ( $\text{CO}_2$ ,  $\text{H}_2\text{Se}$ ,  $\text{COS}$ , etc.), but also organic molecules, such as those formed in high-energy initiated syntheses based on "primordial atmospheres" modeled after that of Jupiter. A gas chromatograph column designed for highly polar gases such as ammonia will be required so that analysis for hydrogen cyanide, formaldehyde, and their oligomers can be performed without penalty.

### 3. Proposed System

To perform the range of analyses that would be required while allowing for multiple determinations as a function of altitude, a dual-column gas chromatograph is proposed; one column would be optimized for polar and the other for non-polar compounds. To achieve the required sensitivity and to minimize pumping requirements, a hydrogen carrier gas with an electrolytic generator-separator will be required. For redundancy, direct inlet via molecular leak to the mass spectrometer and a separate ion cross-section chromatographic detector will be provided. The direct inlet mode would also be required for hydrogen analysis unless an independent chromatograph with a different carrier gas is employed. A possibility would be hydrogen-deuterium-helium analysis on a palladium column with krypton carrier gas. Sampling of an atmosphere over a pressure range of a few millibars to several atmospheres would require multiple samplers of varying volumes. Figure 1 illustrates a proposed atmospheric gas chromatograph-mass spectrometer.

Following capture of a suitably sized sample, the atmospheric gases are swept by a stream of electrolytically generated hydrogen through the polar column and the non-polar column bypass and/or the polar column bypass and the non-polar (molecular sieve) column. This will give

Table 1. Fragment interference with compound determination

Parent Compound $m/e$	$\text{CH}_4$	$\text{NH}_3$	$\text{H}_2\text{O}$	$\text{CO}_2$	$\text{HCN}$	A	Ne
12	C			C	$\text{C}^{12}$		
13	$\text{C}^{13}$ , CH				$\text{C}^{13}$ , CH		
14	$\text{C}^{13}\text{H}$ , $\text{CH}_2$ , CD	N			$\text{N}^{14}$		
15	$\text{CH}_3$ , $\text{C}^{13}\text{H}_2$ , CDH	NH, $\text{N}^{15}$			$\text{N}^{15}$		
16	$\text{CH}_4$ , $\text{C}^{13}\text{H}_3$ , $\text{CH}_2\text{D}$	$\text{NH}_2$ , $\text{N}^{15}\text{H}$ , ND	O	O			
17		$\text{NH}_3$ , $\text{N}^{15}\text{H}_2$ , NHD	OH, $\text{O}^{17}$	$\text{O}^{17}$			
18		$\text{NH}_4$ , $\text{N}^{15}\text{H}_3$ , $\text{NH}_2\text{D}$	$\text{H}_2\text{O}$ , OD, $\text{O}^{17}\text{H}$ , $\text{O}^{18}$	$\text{O}^{18}$		$\text{A}^{36++}$	
19			HDO, $\text{H}_2\text{O}^{17}$ , $\text{O}^{18}\text{H}$ , $\text{H}_3\text{O}$				
20			$\text{H}_2\text{O}^{18}$			$\text{A}^{40++}$	$\text{Ne}^{20}$

satisfactory separations of both non-polar materials, i.e., the inert gases methane, and nitrogen, and polar materials, i.e., ammonia, water, and hydrogen cyanide. At this point, a small amount of a second carrier gas is added to maintain flow after removal of hydrogen in the transmodulator (Ref. 5). A cross-section detector provides redundancy in the event of mass spectrometer failure, giving relative amounts of each gas passing through, and tentative identification based on gas chromatograph retention time. Introduction of components into the mass spectrometer permits positive identification and quantitative determination of isotope ratios.

## References

1. Lewis, J. S., "Observability of Spectroscopically Active Compounds in the Atmosphere of Jupiter," *Icarus*, Vol. 10, p. 393, 1969.
2. Space Science Board, "The Outer Solar System: A Program for Exploration," *Proc. Nat. Acad. Sci. U.S.A.*, Washington, 1969.
3. Sagan, C., "The Production of Organic Molecules in Planetary Atmospheres," *Astron. J.*, Vol. 65, p. 499, 1969.
4. Sagan, C., *et al.*, "Organic Molecules and the Coloration of Jupiter," *Nature*, Vol. 21, p. 273, 1967.
5. Lovelock, J. E., Charlton, K. W., and Simmonds, P. G., "The Palladium Transmodulator: A New Component for the Gas Chromatograph," *Anal. Chem.*, Vol. 41, p. 1048, 1969.

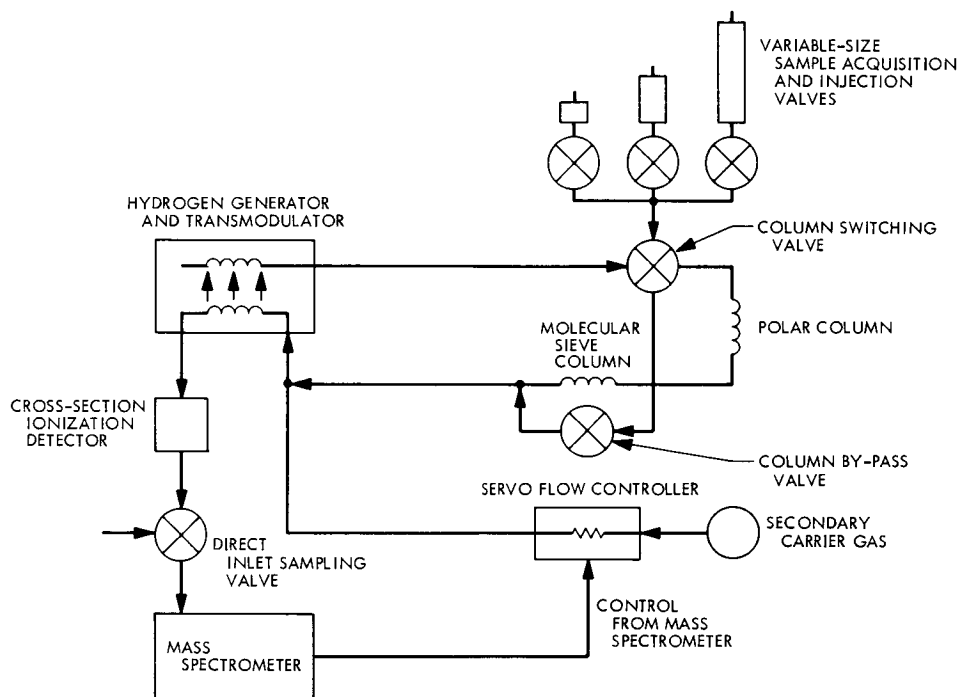


Fig. 1. Atmospheric gas chromatograph-mass spectrometer

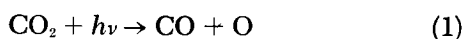
### III. Physics

#### SPACE SCIENCES DIVISION

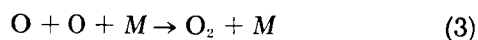
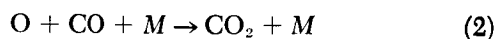
#### A. The Role of CO<sub>2</sub> in the Mars and Venus Atmospheres, W. B. DeMore and C. Dede<sup>1</sup>

##### 1. Introduction

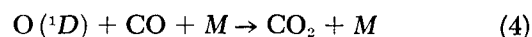
At the present time, there is great interest in the formulation of models that can successfully account for earth-based and *Mariner* spacecraft observations of the atmospheres of Mars and Venus (Refs. 1, 2, and 3).<sup>2</sup> In this connection, some difficulty has been encountered in explaining the rather remarkable fact that the upper regions of these atmospheres consist largely of undissociated CO<sub>2</sub> rather than the expected CO and O. From the known flux of solar radiation  $h\nu$ , the rate of CO<sub>2</sub> dissociation



can be calculated. Atomic oxygen produced in Reaction (1) may recombine with CO or may form O<sub>2</sub>:

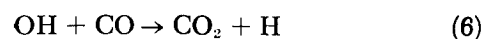


where  $M$  is the third-body. The equilibrium extent of CO<sub>2</sub> dissociation is then determined by the relative rates of dissociation (Reaction 1) and recombination (Reaction 2). The difficulty, however, is that the rate of recombination is too low to account for the low observed level of CO<sub>2</sub> dissociation. This paradox has led to a proposal of several alternative mechanisms to account for the rapid CO<sub>2</sub> recombination. These include a suggestion that Reaction (2) actually involves electronically excited atomic oxygen O(<sup>1</sup>D) rather than the normal ground-state oxygen O(<sup>3</sup>P):



However, despite the fact that CO<sub>2</sub> photolysis does give O(<sup>1</sup>D), it has been shown that O(<sup>1</sup>D) is rapidly quenched to the ground state by both CO<sub>2</sub> (Refs. 4 and 5) and CO (Refs. 5 and 6). Thus, CO<sub>2</sub> cannot recombine by Reaction (4).

The suggestion has also been made (Ref. 7) that traces of H<sub>2</sub> may catalyze the recombination process by the following reactions:

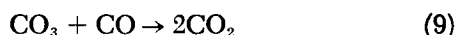


<sup>1</sup>Summer employee at JPL.

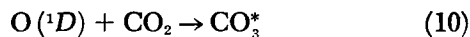
<sup>2</sup>See also a forthcoming paper by M. B. McElroy and D. M. Hunten to be published in the *J. Geophys. Res.*

$$\text{O} + \text{CO} + \text{H}_2 \rightarrow \text{CO}_2 + 2\text{H} \quad (7)$$

The basic problem in each of the above mechanisms is that the rate-determining step is a three-body reaction, which is slow at the low densities of the upper atmosphere. In order to avoid this difficulty, attempts have been made to devise a two-body mechanism for  $\text{CO}_2$  recombination. These usually involve a proposed  $\text{CO}_3$  intermediate, as follows:

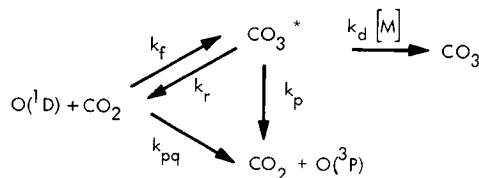


McElroy and Hunten (Footnote 2) have avoided the third-body problem by suggesting that  $\text{CO}_3$  is stabilized by radiative emission:


$$\text{CO}_3^* \rightarrow \text{CO}_3 + h\nu \quad (11)$$

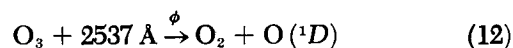
The present article describes experiments designed to measure the rate and pressure dependence of Reaction (8) in the gas phase. This work is an extension of earlier experiments (Ref. 11) on the same reaction in a liquid CO<sub>2</sub> solvent.

**a. Mechanism.** For purposes of discussion it is convenient to write down a detailed mechanism for the  $O(^1D)$ -CO<sub>2</sub> interaction:

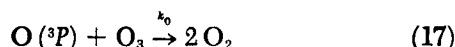
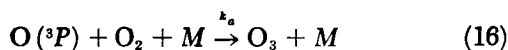
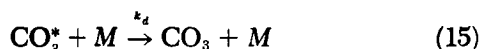
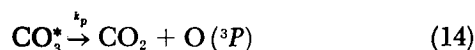
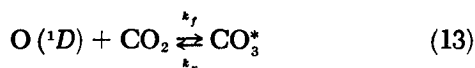


Some information is available from earlier work. In Ref. 11, it was shown that  $k_f$  is comparable to some of the highest known rate constants of  $O(^1D)$ . It was further shown that  $k_f \gg k_{pq}$ , which means that the  $O(^1D)$ - $CO_2$  interaction always involves a  $CO_3^*$  intermediate. Furthermore, this latter point can be supported by comparison with gas phase measurements (Ref. 15). The rate of  $CO_3$  formation in the liquid, measured relative to the  $O(^1D)$ - $C_3H_8$  reaction rate, was found to be equal to the total rate of  $O(^1D)$  consumption by  $CO_2$  in the gas phase, again measured relative to the  $C_3H_8$  reaction. Thus, all  $O(^1D)$  removal by  $CO_2$  can be accounted for by the  $CO_3$ -forming reaction, and there is no evidence for any direct formation of  $O(^3P)$  by physical quenching.

We consider the photolysis of  $O_3$  at 2537 Å in a mixture containing  $CO_2$ ,  $O_2$ ,  $O_3$ , and  $M$  pressurizing gas. The initial step produces  $O(^1D)$ :



By suitable choice of reaction conditions ( $\text{CO}_2$  in great excess over  $\text{O}_2$  and  $\text{O}_3$ , and using He or Ar as the pressurizing gas), only the reaction of  $\text{O}(^1D)$  with  $\text{CO}_2$  need be considered. The following mechanism then applies:



The quantum yield of  $\text{O}_3$  disappearance  $\Phi_{\text{O}_3}$  is given by

$$\frac{\Phi_{\text{O}_3}}{\phi} = \frac{k_d[M]}{k_d[M] + k_p} + \frac{2k_p}{k_d[M] + k_p \left( \frac{1 + k_a[\text{O}_2][M]}{k_0[\text{O}_3]} \right)} \quad (18)$$

Under the conditions of these experiments, the second term on the right-hand side can be shown to be negligible, because  $k_a[\text{O}_2][M] \gg k_0[\text{O}_3]$ , so that

$$\phi/\Phi_{\text{O}_3} = 1 + \frac{k_p}{k_d[M]} \quad (19)$$

Thus, a plot of  $1/\Phi_{\text{O}_3}$  versus  $1/[M]$  yields the ratio  $k_p/k_d$  from the slope. That information, combined with our previous results, makes possible an estimation of the rate constant for  $\text{CO}_3$  formation in the gas phase.

**b. Apparatus and experimental conditions.** The photolyses were carried out in a cylindrical, stainless-steel cell fitted with quartz windows at each end. The optical path length through the cell was 5 cm. The cell was designed to operate at pressures up to 2500 psi. The temperature of the cell was controlled by a flow of chilled  $\text{N}_2$  gas that passed through a coil of copper tubing wrapped around the exterior of the cell. The cell was suspended in an evacuated chamber to facilitate temperature control.

Most of the experiments were conducted at  $-30^\circ\text{C}$ , with reaction mixtures consisting of 65 psi  $\text{CO}_2$ , 50 mm  $\text{O}_2$ , 1–2 mm  $\text{O}_3$ , and sufficient He to bring the total pressure within the experimental range of 100 to 2500 psi. A few experiments were carried out with Ar instead of He as the pressurizing gas.

The mixtures were irradiated with 2537 Å light from a low-pressure Hg lamp, and the light intensity was measured with a calibrated Eppley thermopile. The  $\text{O}_3$  concentration was monitored spectrophotometrically using an extinction coefficient of  $3030 \text{ mole}^{-1} \text{ cm}^{-1}$  at 2537 Å.

### 3. Results

**a. Dependence of  $\Phi_{\text{O}_3}$  on  $M$  pressure.** Table 1 shows the observed values of  $\Phi_{\text{O}_3}$  with either  $\text{CO}_2$  or  $\text{N}_2$  as the substrate. The  $\Phi_{\text{O}_3}$  was always higher with  $\text{CO}_2$  than with  $\text{N}_2$ , and the difference increased with increasing total pressure. No appreciable difference was noted in the values of  $\Phi_{\text{O}_3}$  for experiments with 65 psi  $\text{CO}_2$  and 35 psi  $\text{CO}_2$ . This latter result is expected since no competing path for  $\text{O}(^1D)$  loss should be significant at either  $\text{CO}_2$  pressure. The only possible dependency on  $\text{CO}_2$  pressure would arise through a difference in the efficiencies of  $\text{CO}_2$  and He as third bodies. In this connection, the quantum yields obtained with Ar as pressurizing gas were about the same, or perhaps slightly higher, than those obtained with He.

**Table 1. Effect of pressure on the quantum yields of  $\text{O}_3$  photolysis in the presence of  $\text{CO}_2$  or  $\text{N}_2$**

Total pressure, psi	Pressurizing gas	$\text{O}_2$ pressure, mm	Substrate pressure, psi	$\Phi_{\text{O}_3}$	
				$\text{CO}_2$	$\text{N}_2$
100	He	50	35	0.007	0.004
100	He	50	65	0.013	0.002
250	He	50	35	0.022	0.004
250	He	50	65	0.023	0.002
500	He	50	35	0.037	0.003
500	He	50	65	0.033	0.001
500	He	25	65	0.048	—
500	Ar	50	65	0.040	—
1000	He	50	35	0.064	0.002
1000	He	50	65	0.062	0.005
1000	$\text{N}_2$	50	35	0.013	—
1000	$\text{N}_2$	50	65	0.025	—
1000	He	150	65	0.049	—
1000	He	400	65	0.038	0.007
1000	Ar	0	65	0.059	0.009
1000	Ar	50	65	0.059	—
1000	Ar	80	65	0.065	—
1000	Ar	350	65	0.070	—
1250	He	50	35	0.063	—
1250	He	50	65	0.067	0.006
1250	Ar	50	65	0.077	—
1250	Ar	80	65	0.076	—
2000	He	50	65	0.110	0.008
2000	He	150	65	0.067	—
2500	He	50	35	0.127	—
2500	He	50	65	0.165	—

No dependence of  $\Phi_{O_3}$  on the  $O_2$  partial pressure was found. This is not inconsistent with Eq. (18), which would predict  $\Phi_{O_3} = 2$  at  $[O_2]/[O_3] = 0$ , because the actual ratio of  $O_2$  to  $O_3$  is never exactly zero. Therefore, it can be shown that, at the temperature and pressures of these experiments, very small  $O_2$  concentrations are sufficient to scavenge all  $O(^3P)$  atoms and thus prevent any  $O_3$  loss.

**b. Tests for possible effects of impurities or walls on  $\Phi_{O_3}$ .** Several experiments were carried out to test the possibility that the difference in  $\Phi_{O_3}$  for  $CO_2$  and  $N_2$  was due to some spurious feature of the  $CO_2$  mixtures (such as an impurity in the  $CO_2$ ) rather than to  $CO_3$  formation. Addition of 0.13 mm  $H_2$  to a mixture at 1000 psi produced no significant change in  $\Phi_{O_3}$ . Also, addition of 0.18 mm  $H_2O$  (most of which froze out on the walls upon cooling to  $-30^\circ C$ ) had no effect. It was also found that coating the cell walls with Kel-F grease or with Teflon produced no change. Addition of a magnetically driven stirrer had no effect on  $\Phi_{O_3}$ . Finally, experiments with  $CO_2$  of different grades of purity (Matheson Instrument Grade, Matheson Research Grade, and  $CO_2$  that had been passed over  $CuO$  at  $800^\circ C$ ) all gave the same  $\Phi_{O_3}$ .

**c. Determination of  $k_p/k_d$ .** Figure 1 shows a plot of the data from Table 1 according to Eq. (19). The points fall on a reasonably good straight line, except at the lowest pressure where the errors in  $\Phi_{O_3}$  are relatively large. The

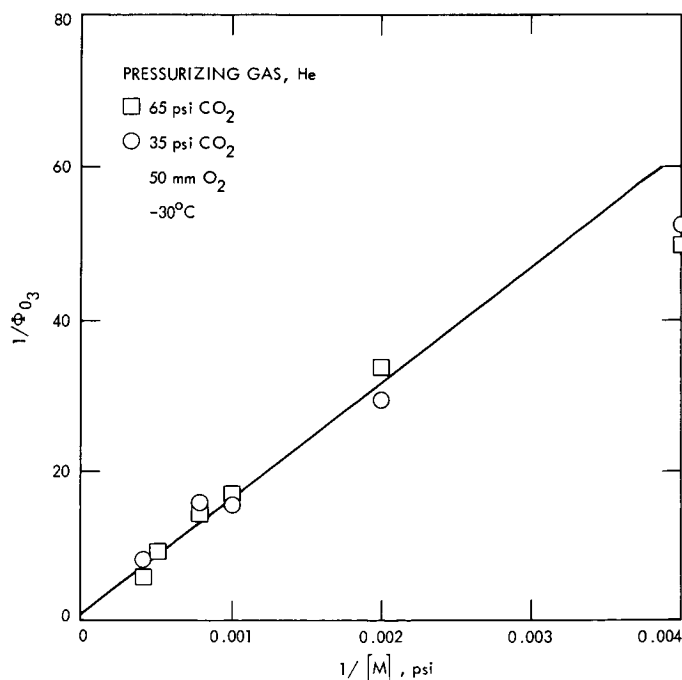


Fig. 1. Plot of data in Table 1 according to Eq. (19)

intercept, while not accurately determined by the data, is clearly consistent with a value near unity, as would be expected if  $\phi = 1$ . From the slope, we find

$$\begin{aligned} k_p/k_d &= 1.6 \times 10^4 \text{ psi} \\ &= 2.6 \times 10^{22} \text{ cm}^{-3} \end{aligned}$$

If we take  $k_d/f = 4.8 \times 10^{-10} \text{ cm}^3 \text{ s}^{-1}$  ( $f$  is the collisional deactivation efficiency, and  $4.8 \times 10^{-10} \text{ cm}^3 \text{ s}^{-1}$  is the calculated rate constant for collisions of He and  $CO_3^*$ , using  $\sigma_{He} = 2.18 \text{ \AA}$  and  $\sigma_{CO_3} = 5.0 \text{ \AA}$ ), we find

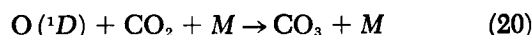
$$k_p = 1.2 \times 10^{13} \text{ s}^{-1} \times f$$

Assuming that  $0.1 > f > 0.01$ , it follows that  $10^{11} < k_p < 10^{12} \text{ s}^{-1}$ .

The overall rate constant of  $CO_3$  formation in the low-pressure limit is  $k = k_f k_d / k_p$ . Taking  $k_f = 2.3 \times 10^{-10} \text{ cm}^3 \text{ s}^{-1}$  (Ref. 5), we find that  $k = 9 \times 10^{-33} \text{ cm}^6 \text{ s}^{-1}$ . The rate constant for quenching of  $O(^1D)$  by  $CO_2$  is  $k_q = (k_f k_p / k_r) + k_p + k_d [M]$  which, at low pressures, reduces to  $k_q = k_f$ . At 1 atm, the quenching reaction is about  $10^3$  faster than  $CO_3$  formation.

#### 4. Discussion

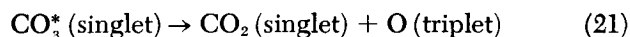
**a. Efficiency of the  $CO_3$ -forming reaction.** There is no reason to doubt that the relatively high rates of  $O_3$  photolysis in the  $CO_2$  experiments are due to the reaction



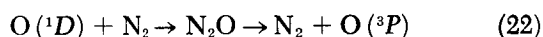
Any other result would have been inconsistent with our previous finding (Ref. 11) that  $CO_3$  is readily formed in liquid  $CO_2$  where the effective concentration of  $M$  is high. At the same time, it is clear that the predissociation reaction giving  $CO_2 + O(^3P)$  instead of  $CO_3$  is the dominant process at ordinary pressures, i.e., less than hundreds of atmospheres. Separate experiments involving detection of  $O(^3P)$  through its reaction with isobutane have confirmed that  $O(^3P)$  does in fact appear as the major product of the  $O(^1D)$ - $CO_2$  reaction at 20 atm.

Our results disagree with those of Clerc and Reiffsteck (Ref. 13) and of Arvis (Ref. 14), who reported efficient  $CO_3$  formation at pressures of 1 atm or less. The reason for the conflicting results is not clear.

It might be thought that the predissociation reaction,

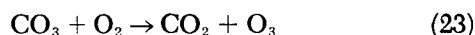


would be slow because it is spin-forbidden. It can be shown, however, that our estimated value of about  $10^{11.5 \pm 1.0} \text{ s}^{-1}$  for  $k_p$  is not unreasonably high, but rather is the expected value. Reactions such as Reaction (21) can be viewed as the unimolecular fission of a vibrationally excited molecule. Processes of this type, involving the splitting of an atom, typically have pre-exponential factors of about  $10^{14 \pm 1} \text{ s}^{-1}$ . Arguments previously given (Ref. 15) in connection with the quenching of  $\text{O}(^1D)$  by  $\text{N}_2$



apply equally well to the present case of  $\text{CO}_2$  and show that the process may be forbidden only by a factor of  $10^{-1}$  or  $10^{-2}$ . Thus, the "forbidden" pre-exponential factor may be as high as  $10^{13 \pm 1} \text{ s}^{-1}$ . Further, the excess vibrational energy of  $\text{CO}_3^*$  permits the rate constant  $k_p$  to approach to within about 1/10 the effective pre-exponential factor, or about  $10^{12 \pm 1} \text{ s}^{-1}$  in this case. This is essentially our measured value of  $k_p$ .

The argument cannot be made that  $\text{CO}_3$  formation in our experiments was masked by secondary reactions such as

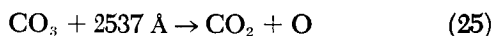


or



(Note that Reaction 24 is different from Reaction 21 in that it does not involve  $\text{CO}_3^*$ .) The reason is that such processes are inconsistent with our observed dependence of  $\Phi_{\text{O}_3}$  on total pressure. Further, Reactions (23) and (24) cannot explain the difference between our results and those of other workers (Refs. 13 and 14) because the same reactions would have been expected under their experimental conditions.

Loss of  $\text{CO}_3$  by photolysis,



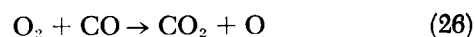
also can be ruled out because most of the 2537 Å light is absorbed by  $\text{O}_3$  and no dependence of  $\Phi_{\text{O}_3}$  on the  $\text{O}_3$  concentration was observed.

#### ***b. Implications regarding the planetary atmospheres.***

The results of this work preclude any role of  $\text{CO}_3$  (formed by Reaction 20) in the planetary atmospheres. Although

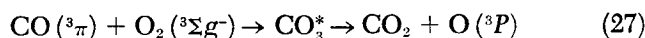
pressures in the lower atmosphere of Venus reach the 100-atm range, no  $\text{CO}_3$  formation will occur because  $\text{O}(^1D)$  is produced photochemically only in the outer reaches of the atmosphere and does not live long enough to diffuse into the high-pressure region. Thus, the question remains as to how  $\text{CO}_2$  is recombined in the planetary atmospheres.

It is unlikely that any chemical mechanism alone can account for the apparently low level of  $\text{CO}_2$  dissociation in the upper atmospheres of Mars and Venus. It is more probable that the explanation lies in the effect of strong mixing, as proposed by Shimizu (Refs. 1 and 2), coupled perhaps with an efficient oxidation of CO at lower altitudes by a bimolecular reaction involving OH (the catalytic reaction). An alternative mechanism would be a direct reaction between  $\text{O}_2$  and CO,



in which either the  $\text{O}_2$  or CO is electronically excited. In this connection, it has been previously shown (Ref. 16) that CO is rapidly oxidized by the  $^3\Sigma u^+$  state of  $\text{O}_2$ , with the reaction in fact proceeding through a triplet  $\text{CO}_3$  intermediate.

Equally effective might be the reaction of triplet  $\text{CO}(^3\pi)$  with  $\text{O}_2$ ,



and it may be significant in this regard that emission from the  $^3\pi$  state of CO (the Cameron bands) was observed as a prominent feature of the Mars UV emission spectrum in the *Mariner* Mars 1969 experiments.

#### **References**

1. Shimizu, M., *Icarus*, Vol. 9, p. 593, 1968.
2. Shimizu, M., *Icarus*, Vol. 10, p. 11, 1969.
3. McElroy, M. B., *J. Geophys. Res.*, Vol. 73, p. 1513, 1968.
4. Preston, K. F., and Cvetanovic, R. J., *J. Chem. Phys.*, Vol. 45, p. 2888, 1966.
5. Young, R. A., Black, G., and Slinger, T. G., *J. Chem. Phys.*, Vol. 49, p. 4758, 1968.
6. Raper, O. F., and DeMore, W. B., *J. Chem. Phys.*, Vol. 40, p. 1053, 1964.
7. Reeves, R. R., Jr., et al., *J. Phys. Chem.*, Vol. 70, p. 1637, 1966.
8. Donahue, T. M., *J. Atmos. Sci.*, Vol. 25, p. 568, 1968.
9. Moll, N. G., Clutter, D. R., and Thompson, W. E., *J. Chem. Phys.*, Vol. 45, p. 4469, 1966.

10. Weissberger, E., Breckenridge, W. H., and Taube, H., *J. Chem. Phys.*, Vol. 47, p. 1764, 1967.
11. DeMore, W. B., and Jacobsen, C. W., *J. Phys. Chem.*, Vol. 73, p. 2935, 1969.
12. Slanger, T. G., *J. Chem. Phys.*, Vol. 45, p. 4127, 1966.
13. Clerc, M., and Reiffsteck, A., *J. Chem. Phys.*, Vol. 48, p. 2799, 1968.
14. Arvis, M., *J. Chem. Phys.*, Vol. 66, p. 517, 1969.
15. DeMore, W. B., and Raper, O. F., *Astrophys. J.*, Vol. 139, p. 1381, 1964.
16. Raper, O. F., and DeMore, W. B., *J. Chem. Phys.*, Vol. 40, p. 1047, 1964.

## B. Gage Transformation of the Second Post-Newtonian Equations of Hydrodynamics in General Relativity, F. B. Estabrook, B. K. Harrison, and H. D. Wahlquist

### 1. Introduction

In Ref. 1, S. Chandrasekhar and Y. Nutku have derived the second post-Newtonian equation of hydrodynamics in an analytically very straightforward way. There are, nevertheless, several delicate remaining questions of covariant interpretation, in particular of the gage transformation that they find. In this article, we will discuss these issues from the uniform standpoint of *active* transformation, which is to say, the simultaneous "dragging" over the coordinate manifold of all geometric objects, along the congruence generated by an arbitrary vector field  $V^i$ . For infinitesimal dragging transformations, we normalize  $V^i$  so that from any geometric object  $f$  (indices suppressed) we obtain the analytically new, or dragged, object  $f^*$  according to

$$f^* = f + \mathcal{L}_{V^i} f$$

where  $\mathcal{L}_{V^i}$  is the Lie derivative with respect to  $V^i$ . For example, a scalar  $f$  becomes  $f^* = f + f_{,i} V^i$ ; a second-order covariant tensor  $g_{ij}$  becomes

$$g_{ij}^* = g_{ij} + g_{ij,k} V^k + V^k_{,i} g_{kj} + V^k_{,j} g_{ik}$$

Any tensor equation is preserved under such a dragging:  $f = 0$  implies  $f^* = 0$ . Indeed, any *intrinsic* description or relation characterizing the fields on the manifold is left unchanged by the dragging and it is, in fact, simply equivalent to arbitrary change of coordinatization of the points of the manifold.

### 2. Gage Transformation

Conforming to the notation and definitions of Ref. 1, we first write the second post-Newtonian metric as

$$\left. \begin{aligned} g_{00} &= 1 - \frac{2U}{c^2} + \frac{2}{c^4} (U^2 - 2\Phi) + \frac{1}{c^6} Q_{00} \\ g_{0\alpha} &= \frac{1}{c^3} P_\alpha + \frac{1}{c^5} Q_{0\alpha} \\ g_{\alpha\beta} &= - \left( 1 + \frac{2U}{c^2} \right) \delta_{\alpha\beta} + \frac{1}{c^4} Q_{\alpha\beta} \end{aligned} \right\} \quad (1)$$

We note that, in the adopted coordinate system,  $U$  is a non-local, non-tensorial function of the mass density scalar,  $\rho$ , viz.,

$$U_{,\alpha\alpha} = -\rho \quad \text{or} \quad U = \int \frac{\rho}{r} d^3x \quad (2)$$

Similarly, non-covariant definitions obtain for the quantities (Ref. 1)  $\phi$ ,  $\Phi$ ,  $U_\alpha$ , and  $P_\alpha$ . The dragging rule for such quantities is not clear and, in fact, will not be used. On the other hand, the energy momentum tensor  $T_{ij}$  obtained from Expression (1) and the field equations is (indeed, required to be)

$$T_{ij} = \rho (c^2 + \pi + p/\rho) u_i u_j - p g_{ij} \quad (3)$$

and so the scalars,  $\rho$ ,  $\pi$ , pressure  $p$  and 4-velocity

$$u^i = \left( 1 + \frac{1}{2} \frac{v^2}{c^2} + \frac{U}{c^2}, \frac{v_\alpha}{c} \right)$$

are covariantly defined, and so will be dragged. The equations of motion for these, in the second post-Newtonian theory, follow from

$$T^j_{i,j} = 0 \quad (4)$$

We now define an arbitrary 4-vector field  $V^i$ , formed to the correct orders in  $1/c$  from arbitrary components  $W$  and  $W_\alpha$ , according to

$$V^i = \left( \frac{1}{c^5} W, -\frac{1}{c^4} W_\alpha \right) \quad (5)$$

We will show that second post-Newtonian gage transformation consists of dragging all tensor entities, and

tensorial equations, along the  $V^i$  congruence. We first write all the active transformations of tensor quantities:

$$\left. \begin{aligned} g_{00}^* &= g_{00} + \frac{2}{c^6} W_\alpha \frac{\partial U}{\partial x^\alpha} + \frac{2}{c^6} \frac{\partial W}{\partial t} \\ g_{0\alpha}^* &= g_{0\alpha} + \frac{1}{c^5} \frac{\partial W_\alpha}{\partial t} + \frac{1}{c^5} \frac{\partial W}{\partial x^\alpha} \\ g_{\alpha\beta}^* &= g_{\alpha\beta} + \frac{1}{c^4} \left( \frac{\partial W_\alpha}{\partial x^\beta} + \frac{\partial W_\beta}{\partial x^\alpha} \right) \end{aligned} \right\} \quad (6)$$

$$\left. \begin{aligned} \rho^* &= \rho - \frac{1}{c^4} W_\alpha \frac{\partial \rho}{\partial x^\alpha} \\ p^* &= p - \frac{1}{c^4} W_\alpha \frac{\partial p}{\partial x^\alpha} \end{aligned} \right\} \quad (7)$$

$$v_\alpha^* = v_\alpha - \frac{1}{c^4} W_\beta \frac{\partial v_\alpha}{\partial x^\beta} + \frac{1}{c^4} v_\beta \frac{\partial W_\alpha}{\partial x^\beta} + \frac{1}{c^4} \frac{\partial W_\alpha}{\partial t} \quad (8)$$

$$\left. \begin{aligned} T_{ij}^* &= \rho^* (c^2 + \pi^* + p^*/\rho^*) u_i^* u_j^* - p^* g_{ij}^* \\ T_{i,j}^* &= 0 \text{ (covariant differentiation using } g_{ij}^*) \end{aligned} \right\} \quad (9)$$

To conform with the notation of Ref. 1, we now also define new quantities  $U^*$  and  $Q_{ij}^*$  by setting

$$U_{,\alpha\alpha}^* = -\rho^* \quad \text{or} \quad U^* = \int \frac{\rho^*}{r} d^3x \quad (10)$$

and

$$\left. \begin{aligned} g_{00}^* &= 1 - \frac{2U^*}{c^2} + \frac{2}{c^4} (U^{*2} - 2\Phi^*) + \frac{1}{c^6} Q_{00}^* \\ g_{0\alpha}^* &= \frac{1}{c^3} P_\alpha^* + \frac{1}{c^5} Q_{0\alpha}^* \\ g_{\alpha\beta}^* &= - \left( 1 + \frac{2U^*}{c^2} \right) \delta_{\alpha\beta} + \frac{1}{c^4} Q_{\alpha\beta}^* \end{aligned} \right\} \quad (11)$$

By comparing Eqs. (6) and (11), there results precisely the relations between  $Q_{ij}$  and  $Q_{ij}^*$  previously given in Ref. 1. In particular, equating the  $g_{00}^*$  of Eqs. (6) and (11) gives

$$2W_\alpha \frac{\partial U}{\partial x^\alpha} + 2 \int \frac{\partial W}{\partial t} + Q_{00} = 2 \int \frac{W_\alpha \frac{\partial \rho}{\partial x^\alpha}}{r} d^3x + Q_{00}^*$$

or

$$Q_{00} + 2W_0 + 2 \frac{\partial W}{\partial t} = Q_{00}^*$$

where  $W_0$  satisfies

$$\nabla^2 W_0 = \nabla^2 W_\alpha \frac{\partial U}{\partial x^\alpha} + 2 \frac{\partial W_\alpha}{\partial x^\beta} \frac{\partial^2 U}{\partial x^\alpha \partial x^\beta}$$

After making such a transformation, it is clear that if we drop the stars in Eqs. (9), (10), and (11), we formally recover Eqs. (1), (2), (3), and (4). In any event, as mentioned previously, the intrinsic nature of any second post-Newtonian solution is unchanged by dragging; by the definition of  $U^*$  and  $Q_{ij}^*$  the form of the dragged metric as a function of  $\rho^*$ ,  $p^*$ , and  $v_\alpha^*$  is also kept invariant. But  $\rho^*$ ,  $p^*$ ,  $v_\alpha^*$ ,  $Q_{ij}^*$  are all now *analytically different functions of the independent variables*  $x^\alpha, t$ .

### 3. Gage Choice

Chandrasekhar and Nutku (Ref. 1) do not clearly state that  $\rho$ ,  $p$ , and  $v_\alpha$  have to be dragged *at the same time* that the  $Q_{ij}$ 's are, although this is certainly required for understanding their statement following the equations of motion (Ref. 1, Eq. 54) that these equations are gage invariant. If gage transformation of the metric (Eq. 1) were taken to mean *only* transformation of the  $Q_{ij}$ 's in a given second post-Newtonian solution, then the resulting new metric is certainly *intrinsically* and *metrically* changed, and will in general *not* be a second post-Newtonian solution.

On the other hand, we may begin at some epoch—on an initial surface  $t = t_0 = \text{constant}$  with an initial distribution of  $\rho$ ,  $p$ ,  $v_\alpha$ , and  $Q_{ij}$ . We may consider separately two integrations for the “self-consistent” time development of these quantities, using the Ref. 1 equations of motion for  $\rho$ ,  $p$ ,  $v_\alpha$ , and the Ref. 1 spatial field equations (Eqs. 23, 28, and 33) for the  $Q_{ij}$ 's. For the first integration, we make throughout the “gage choice”  $W = 0$ ,  $W_\alpha = 0$ ; or for the second, we consider how the quantities self-consistently develop in time with some other gage choice, still having  $W$ ,  $W_\alpha$  zero at  $t_0$ , but chosen arbitrarily thereafter. The dynamical equations of motion for  $\rho$ ,  $p$ ,  $v_\alpha$  explicitly involve the  $Q_{ij}$ , but as we have remarked, these equations are properly gage invariant; and so in the *second* case considered, the initial  $\rho$ ,  $p$ ,  $v_\alpha$ ,  $Q_{ij}$  must develop precisely into the analytically different fields  $\rho^*$ ,  $p^*$ ,  $v_\alpha^*$  and  $Q_{ij}^*$  (which intrinsically *are* the same solution as in

the first case). It is perhaps from this dynamical viewpoint that one is tempted to think of the gage transformation of Ref. 1 as only being required to be applied to the field quantities  $Q_{ij}$  through the Laplacian equations which determine these, and not explicitly to the fluid variables  $\rho$ ,  $p$ ,  $v_a$ , as these last would then "automatically" be changed in the analytical result (the complete time history and the metric).

We have clearly seen the four undetermined gage functions as expressing coordinate indeterminacy in integration of the equations of motion. The indeterminacy did not arise in previous work on the first post-Newtonian equations because the *a priori* choice was made that the  $1/c^2$  term in  $g_{\alpha\beta}$  be proportional to  $\delta_{\alpha\beta}$ . The only indeterminacy remaining there was in the time coordinatization: indeed, purely time-like gage transformation, in this case using 4-vector field

$$V^i = \left( \frac{1}{c^3} W, 0, 0, 0 \right)$$

is conceptually especially simple, since scalars like  $\rho$  and  $p$ , and the timelike unit vector  $v^i$ , are not changed at all in the appropriate order. Timelike dragging appears solely as a change of gage in expressing the field quantities  $h_{00}$  and  $h_{0a}$ , in especially close analogy to the gage transformation of electromagnetism. As an example, if  $W = (\partial X / \partial t) / 2$ , the post-Newtonian metric whose analytical form satisfies the Landau-Lifschitz gage condition is changed to one in the so-called radiation gage. We should emphasize, however, that a general dragging could equally be performed at the post-Newtonian level; the spatial metric would then no longer be conformally flat.

#### Reference

1. Chandrasekhar, S., and Nutku, Y., "The Second Post-Newtonian Equations of Hydrodynamics in General Relativity," *Astrophys. J.*, Vol. 158, pp. 55-79, 1969.

### C. An Ion Cyclotron Resonance Study of the Ion-Molecule Reactions in Methane-Ammonia Mixtures, W. T. Huntress, Jr., and D. D. Elleman

#### 1. Introduction

Ion-molecule reactions occurring in mixtures of methane and ammonia may be important processes for chemical synthesis in radiolytic systems. Recent work on electrical

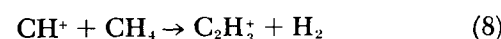
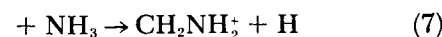
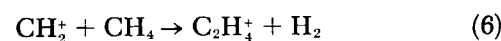
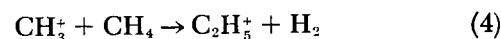
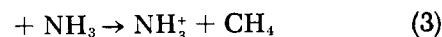
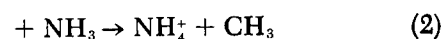
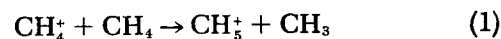
discharges in methane and ammonia have shown that large yields of unsaturated hydrocarbons (Refs. 1 and 2), HCN (Refs. 1, 2, and 3), and  $\alpha$ -aminonitriles (Refs. 2 and 3) are obtained. Mixtures of methane, ammonia, and water subjected to various forms of irradiation such as electrical discharges, ultraviolet light, and high-energy electrons yield significant amounts of many amino acids (Refs. 4 and 5). The DNA bases adenine and guanine have been identified in some of these experiments, and the principal intermediate for these products is believed to be hydrogen cyanide produced from methane and ammonia (Refs. 4 and 5). These experiments have fascinating biological implications in regard to the origin of life, and are particularly pertinent to the study of the chemistry in the atmosphere of the Jovian planets. Both methane and ammonia have been clearly identified in the atmosphere of Jupiter, and the vivid color-banded atmosphere and Great Red Spot of Jupiter may be due in part to radiation-induced organic synthesis (Ref. 6).

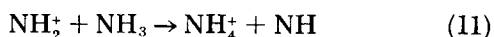
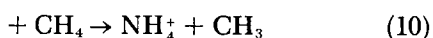
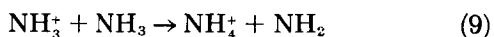
This article describes the results of a study of the ion chemistry in methane-ammonia mixtures using ion cyclotron resonance techniques (SPS 37-53, Vol. III, pp. 149-151 and SPS 37-46, Vol. IV, pp. 205-208). A condensation sequence is observed between the methyl cation,  $\text{CH}_3^+$ , and ammonia molecules which leads to the eventual synthesis of hydrogen cyanide in the methane-ammonia system. The synthesis of hydrogen cyanide in methane-ammonia mixtures has previously been somewhat of a mystery.

#### 2. Experimental Results

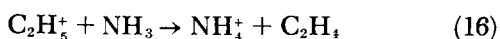
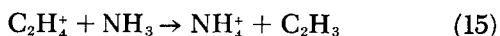
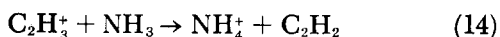
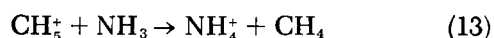
The reactions identified by double-resonance experiments are summarized as follows:

##### Primary ions





### Secondary ions



Of particular interest is the condensation reaction (Reaction 5). The single resonance spectrum in Fig. 2 shows that there are two major condensation products in the methane-ammonia system,  $\text{C}_2\text{H}_5^+$  at  $m/e = 29$  and  $\text{CH}_2\text{NH}_2^+$  at  $m/e = 30$ . Both of these product ions originate from the methyl cation,  $\text{CH}_3^+$ , via Reactions (4) and (5). The  $\text{CH}_2\text{NH}_2^+$  ion at  $m/e = 30$  is the only ion observed in the single resonance spectrum in methane-ammonia mixtures that is not present in the spectrum of either pure gas, and is the only condensation product observed in which a carbon-nitrogen bond is formed.

The proton affinity of ammonia is very high and ammonia is a very effective proton scavenger. At elevated pressures, the  $\text{NH}_4^+$  ion is the only ion observed in the spectrum with the exception of a much smaller peak due to the  $\text{CH}_2\text{NH}_2^+$  ion. This latter ion undergoes proton transfer to ammonia at lower pressures, but at high pressures the proton transfer reaction is no longer observed and the  $m/e = 30$  ion persists in the methane-ammonia single resonance spectrum. This behavior is illustrated

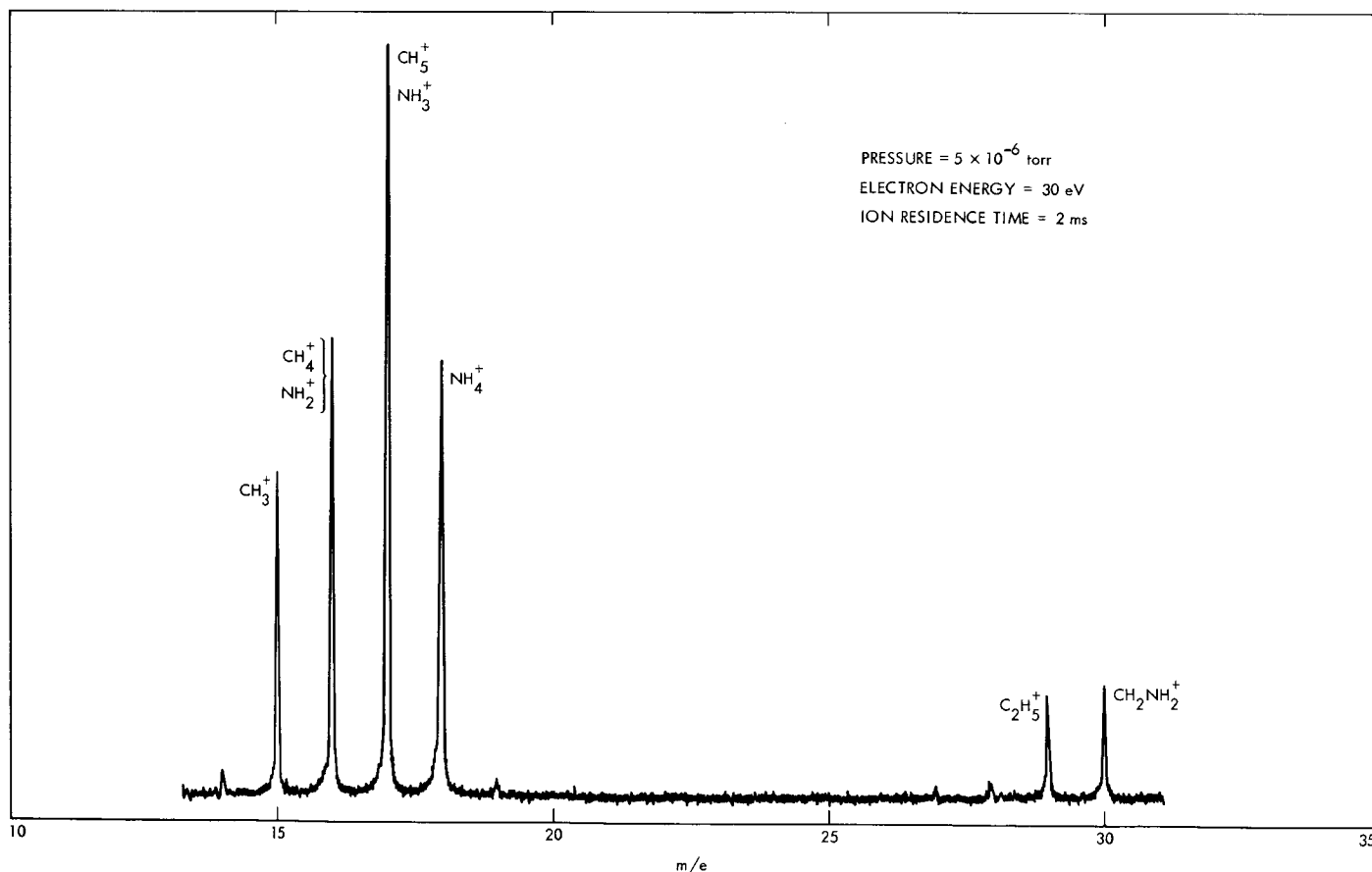


Fig. 2. Single-resonance spectrum of the ions in a 2:1 mixture of methane and ammonia

in Fig. 3a where the other major condensation product, the  $C_2H_5^+$  ion at  $m/e = 29$ , is included for comparison. The proton transfer reaction from  $C_2H_5^+$  to  $NH_3$  proceeds to completion. It is quite apparent from the intensity versus pressure curve for the  $m/e = 30$  ion in Fig. 3a that unlike the  $m/e = 29$  ion, the proton transfer reaction does not proceed to completion and that the reaction is inhibited at higher pressures. This suggests that the ion is initially formed in an excited state, in which the proton transfer reaction is exothermic, and that, at higher pressures, the excited ion is collisionally deactivated to the ground-state ion for which the proton transfer reaction is evidently endothermic. This hypothesis is supported by the double-resonance data given in Fig. 3b.

Figure 3b shows the double-resonance intensity versus pressure for the proton transfer reactions of  $C_2H_5^+$  and  $CH_2NH_2^+$  (Reactions 16 and 17). The double-resonance intensity versus pressure for Reaction (16) parallels the single-resonance intensity almost exactly and shows a decrease in product ion intensity with increasing kinetic energy of the reactant ion typical of exothermic proton transfer reactions. In order to get a negative response, or *decrease* in ion intensity, a necessary condition is that the reaction proceed at thermal ion kinetic energies. The fact that intensity versus pressure curves for the double and single resonance exhibit the exact same behavior with pressure indicates that Reaction (16) proceeds with the same rate at all pressures.

The initial behavior of the double-resonance intensity versus pressure curve in Fig. 3b for Reaction (17) at low pressures is similar and the reaction appears to be exothermic. However, there is a marked deviation with increasing pressure, and the double-resonance response changes sign to become positive. At high pressures, the magnitude of the response for Reaction (17) reaches a plateau, corresponding to the behavior of the single-resonance intensity for the  $CH_2NH_2^+$  ion. The positive response indicates that the product ion intensity increases with increasing reactant ion kinetic energy, which is necessarily the case for an endothermic reaction.

At high pressures, the proton transfer reaction (Reaction 17) is endothermic and this endothermicity is overcome by increasing the translational energy of the reactant ion. The change in reactivity observed for the  $CH_2NH_2^+$  can only be consistent with some change in the state of the ion at high pressures. Reaction (5) is highly exothermic, releasing 74 kcal/mole of energy. Therefore, it is not unreasonable that the initial product should contain a significant amount of internal energy. The rate for Reaction (5),

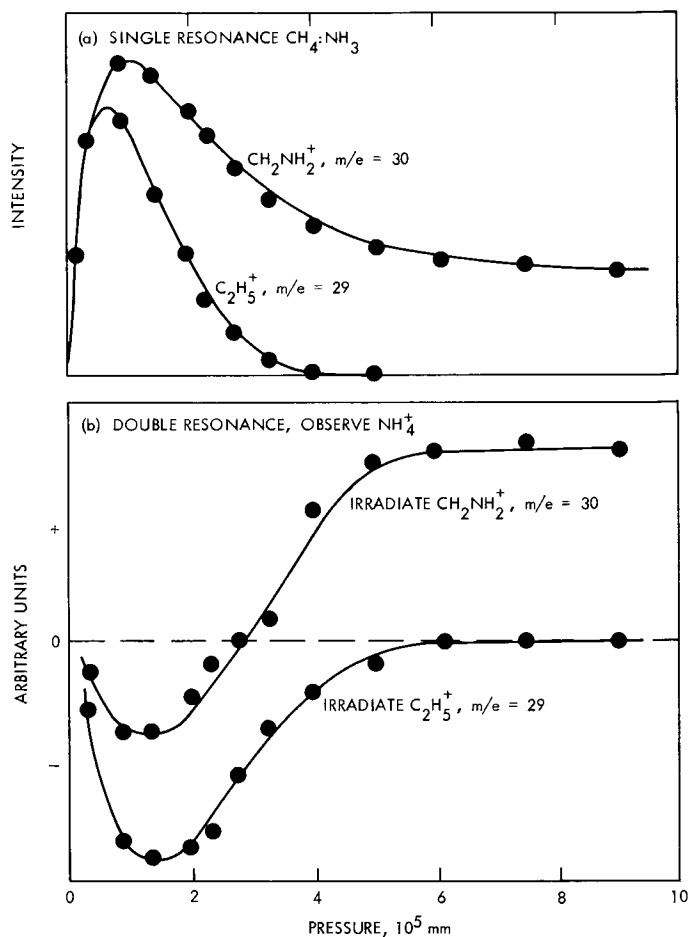


Fig. 3. Single- and double-resonance intensities versus pressure for proton transfer reactions

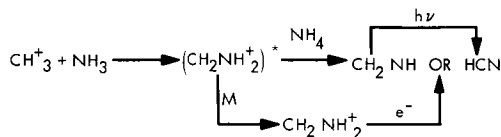
tion (5), determined by the method of Bowers, et al. (SPS 37-54, Vol. III, pp. 158-161) is very fast:  $2 \times 10^{-9}$   $cm^3/molecules \cdot s$ .

### 3. Discussion

Several processes identified in this study are of particular importance to chemical synthesis in discharges and related systems. The major secondary condensation product in methane is the  $C_2H_5^+$  ion. This ion and the  $C_2H_4^+$  ion are inert towards further reaction with methane. The introduction of ammonia, however, provides a very effective proton scavenger and all secondary ions in methane react to produce  $NH_4^+$ , yielding both stable and unstable reactive neutral products. The methyl radical,  $CH_3$ , is produced by Reactions (1), (2), and (10), and the vinyl radical,  $C_2H_3$ , by Reaction (15). Two stable species, ethylene and acetylene, are products of proton transfer from  $C_2H_5^+$  and  $C_2H_3^+$ . Ethylene and acetylene are produced

in significant yields by discharges in methane and ammonia (Refs. 1 and 2), and both molecules readily polymerize under radiolysis.

The production of  $\text{CH}_2=\text{NH}_2^+$  ions by rapid ion-molecule reaction may prove of substantial consequence regarding the synthesis of carbon-nitrogen compounds in discharges and in radiolytic systems containing methane and ammonia. The excited ion will proton-transfer to ammonia producing either methylenimine or hydrogen cyanide. Dissociative recombination of the stabilized ground-state ion with electrons probably leads to formation of methylenimine and possibly hydrogen cyanide.



where  $h\nu$  is the solar radiation. These reactions should all be very fast, the condensation reaction and proton transfer reactions having rates on the order of  $10^{-9}$  cm<sup>3</sup>/molecule-s.

Methylenimine is a highly reactive molecule and has been previously detected only by photolysis studies in rare gas matrices at low temperatures (Ref. 7). Methylenimine is probably readily dissociated into hydrogen cyanide and hydrogen by low-energy photons. Hydrogen cyanide, and also methylenimine, are likely to be important intermediates in the formation of organic molecules of considerable biological interest.

## References

1. Mathews, C. N., and Moser, R. E., "Prebiological Protein Synthesis," *Proc. Nat. Acad. Sci. (U.S.)*, Vol. 56, p. 1087, 1966.
2. Sagan, C., "Molecular Synthesis in Simulated Reducing Planetary Atmospheres," *Astron. J.*, Vol. 65, p. 499, 1960.
3. Ponnamperuma, C., and Woeller, F., " $\alpha$ -Aminonitriles Formed by an Electric Discharge Through a Mixture of Anhydrous Methane and Ammonia," *Curr. Mod. Biol.*, Vol. 1, p. 156, 1967.
4. Ponnamperuma, C., and Gabel, N. W., "Current Status of Chemical Studies on the Origin of Life," *Space Life Sci.*, Vol. 1, p. 64, 1968.
5. "Abiological Synthesis of Some Nuclei Acid Constituents," in *The Origin of Prebiological Systems*. Edited by S. W. Fox. The Academic Press, New York, 1965.
6. Sagan, C., "Jovian Atmosphere: Near Ultraviolet Absorption Features," *Science*, Vol. 159, p. 448, 1968.
7. Moore, C. B., Pimentel, G. C., and Goldfarb, T. D., "Matrix Photolysis Products of Diazomethane: Methylenimine and Hydrogen Cyanide," *J. Chem. Phys.*, Vol. 43, p. 63, 1965.

#### D. Solar Plasma Experiment for Apollo,

C. W. Snyder, D. R. Clay, and M. Neugebauer

## 1. Introduction

A solar wind spectrometer (SWS) is one of the scientific instruments included in the *Apollo* Lunar Surface Experiments Package (ALSEP). The ALSEP system is designed for self-sufficient operation on the lunar surface, collecting and transmitting data to the earth for approximately 2 yr. ALSEP 1 was deployed on the moon by the crew of *Apollo 12* on November 19, 1969, near the *Surveyor III* landing site.

The subsystems of ALSEP 1 include:

- (1) A SNAP-27 radioisotope thermoelectric generator producing about 74 W of continuous power at 16 Vdc.
- (2) A central station that provides a two-way radio link to earth, power conditioning and switching for instrument turn on, data handling and conditioning (including telemetry commutation and control as well as analog-to-digital conversion for certain instruments), and command decoding and signal distribution.
- (3) A seismometer to measure the natural seismicity of the moon, lunar free oscillations, tidal deformations and seismic response to man-made events such as the lunar module lunar impact.
- (4) A magnetometer to measure the time variations and gradients of the local magnetic field.
- (5) An atmosphere detector to measure the pressure of the neutral atmosphere as well as the flux, composition, and velocity of low-energy (suprathermal) positive ions.
- (6) A SWS to measure the temporal variations of the basic properties of the plasma from the sun. A detailed description of this instrument is presented in this article.

Each scientific instrument is emplaced separately on the lunar surface, connected to telemetry and power via a simple cable, and is responsible for its own thermal control and lunar surface mount.

## 2. Experiment Objectives

The SWS was designed for ALSEP with the objective of detecting whatever solar plasma might strike the surface of the moon. Thus, the instrument was required to

be sensitive enough to detect the normal solar-wind flux at any angle in the lunar sky and to measure enough of the properties of the bulk solar wind to establish the nature of its interaction with the moon. At the time the experiment was proposed (1966), nothing whatever was known about this interaction. Subsequently, measurements of the solar wind and the magnetic field were made near the moon by the lunar orbiter *Explorer XXXV* (Refs. 1-6) and rare gases from the solar wind were detected at the lunar surface by the *Apollo 11* Solar Wind Composition Experiment (Ref. 7). The results of both these investigations imply that the solar wind strikes the moon and that little if any change in its properties occur before impact.

The scientific objectives of the SWS are as follows:

- (1) **Existence of solar wind plasma on the moon.** By comparison of solar wind properties measured at the lunar surface with those measured in space near the moon, to determine whether the moon has any effect on the solar plasma other than simply absorbing it.
- (2) **Properties of lunar surface and interior.** If any subtle effects of the moon on solar wind properties are discovered, it may be possible to relate these to properties of the moon itself, such as its magnetic field, its electrical conductivity, the possibility of retaining an atmosphere, or the possible effect of solar corpuscular radiation on the lunar surface layer by the mechanism of sputtering or electrical charging.
- (3) **General solar wind properties.** The motion of waves or discontinuities in the solar wind can be studied by measuring the time intervals between the observations of changes in plasma properties at the moon and at the earth.
- (4) **Magnetospheric tail of earth.** The measurements (during 4 or 5 days around the time of full moon) may permit inferences as to the length, breadth, and structure of the magnetospheric tail of the earth.

### 3. Detection Method

The basic sensor in the SWS is a Faraday cup, which measures the charged particle flux entering the cup by collecting these ions and using a sensitive current amplifier to determine the resultant current flow. Energy spectra of positive and negative charged particles are obtained by applying fixed sequences of retarding potentials to a modulator grid and measuring the resulting changes in current. A squarewave ac voltage is superimposed on a dc voltage to form the retarding potential. This type of detec-

tor measures the flux density of charged particles that enter the aperture and whose energy per unit charge for the component of velocity normal to the cup is within the range corresponding to the upper and lower voltage on the modulator grid. Similar detectors have been flown on a variety of space probes (Ref. 8).<sup>3</sup>

To be sensitive to solar wind plasma from any direction (above the horizon of the moon) and to ascertain its angular distribution, the SWS has an array of seven cups. Since the cups are identical, an isotropic flux of particles would produce equal currents in each cup. If the flux is not isotropic but appears in more than one cup, analysis of the relative amounts of current in the collectors can provide information on the direction of plasma flow and its anisotropy. The central cup faces the vertical, and the remaining six surround it symmetrically, each facing 60 deg off the vertical. The combined acceptance cones of all cups cover most of the upward hemisphere. Each cup has a circular opening, five grids, and a circular collector. The function of the grid structures is to apply an ac modulating field to incoming particles and to screen the modulating field from the inputs to the sensitive preamplifiers (Fig. 4). The entrance apertures of the cups were protected from damage or dust by covers that remained in place until after the departure of the lunar module.

### 4. Electronics Description

The electronics for the instrument is in a temperature-controlled container that hangs below the sensor assembly (a block diagram of the electronics is given in Fig. 5).

The digital circuit programmer performs internal data handling and processing and generates internal commands

<sup>3</sup>See also Vasyliunas, V. M., "Deep Space Plasma Measurements" (to be published in *Methods Exp. Phys.*, Vol. 9).

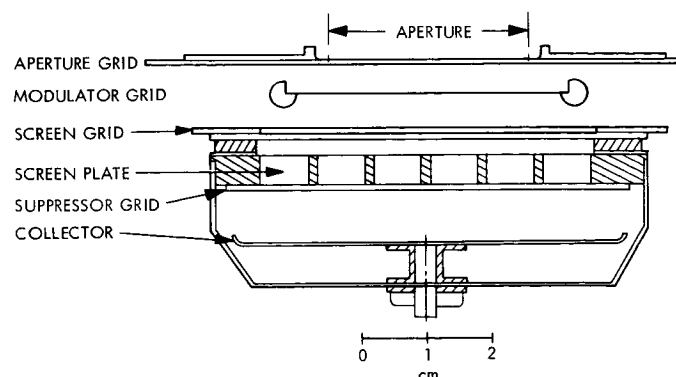


Fig. 4. Faraday cup sensor

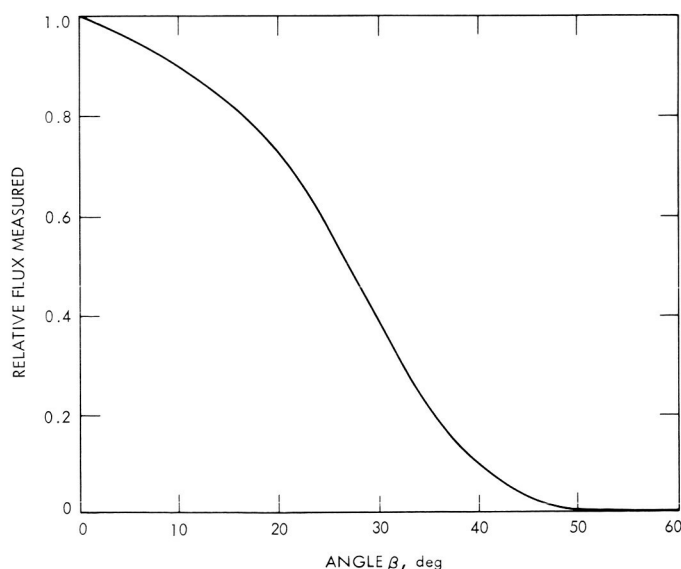


The instrument operates in an invariable sequence in which a complete set of plasma measurements is made every 28.1 s. It consists of 14 energy steps spaced a factor  $2^{1/2}$  apart for positive ions and 7 energy steps spaced a factor 2 apart for electrons. A ground command for high gain modulation increases the ac and dc voltages for all steps by a factor of  $2^{3/4}$ , thus extending the spectrometer energy range. A large number of internal calibrations are provided, and every critical voltage is read out at intervals of 7.5 min or less. The fabrication and a portion of the development and testing of the SWS were conducted for JPL by Electro-Optical Systems (EOS) of Pasadena.

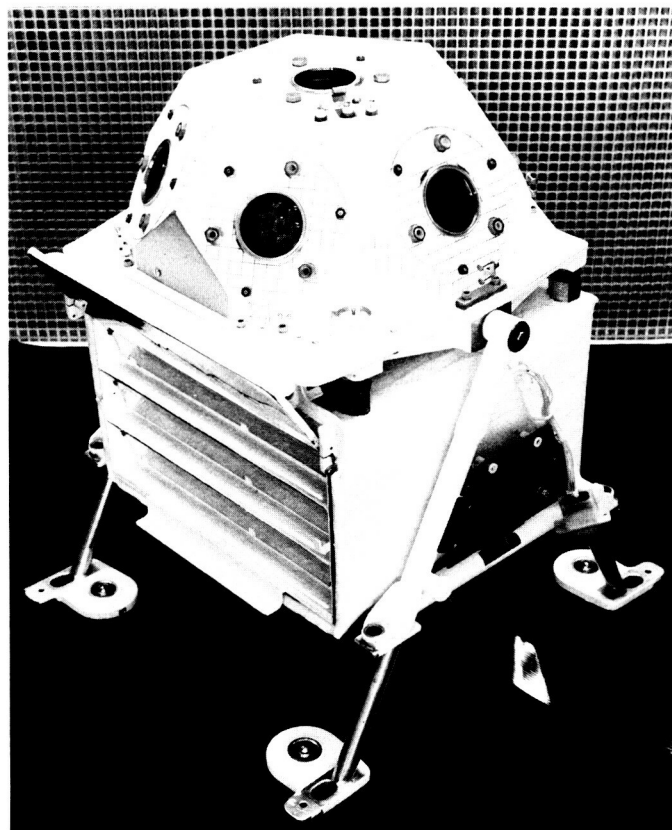
The laboratory plasma calibration of this instrument was conducted by JPL at EOS facilities. The angular dependence of the Faraday cup sensor to plasma was empirically determined and the average for the seven cups agrees quite well with the measured optical transparency. The relative currents measured as a function of the angle (angle  $\beta$ ) between the normal to the cup entrance and the velocity of the bulk laboratory plasma is shown in Fig. 6.

### 5. Instrument Deployment

The SWS was deployed without difficulty, and is shown in Fig. 7 in deployed configuration. Examination of the photograph taken after the SWS was deployed indicates, from the pattern of the shadow on the radiator surfaces (left side of instrument), that the east-west axis of the instrument is actually aligned approximately 2.8 deg north



**Fig. 6. Average angular response of Faraday cup**



**Fig. 7. SWS as deployed on lunar surface**

of east. On the moon, the instrument is hung from a pair of knife edges so that it is free to swing about an east-west horizontal axis, and hence is self-leveling in one dimension. Level about the north-south axis is indicated by a sun sensor that indicates the time that the sun is 30 deg east of the axis of the central cup. The data from the sun sensor indicates that the axis is off-level by about 2.5 deg, the west edge being low. Both these values are well within specified limits. Level about the east-west axis cannot be determined from available photographic information, but should be assured by the self-leveling suspension.

Shortly after deployment, the SWS was turned on to provide background data with sensor covers in place. About 1 h after lunar-module ascent, the covers were removed by command from earth and detection of solar plasma began.

### 6. Instrument Performance

The SWS has functioned as expected during the first month. All subsystem functions have performed properly.

The thermal control has proven adequate with electronics temperatures between  $-16$  and  $+63^{\circ}\text{C}$ , leaving comfortable margins for possible future degradation. Temperature extremes inside the sensor assembly were  $-134$  and  $+63^{\circ}\text{C}$ .

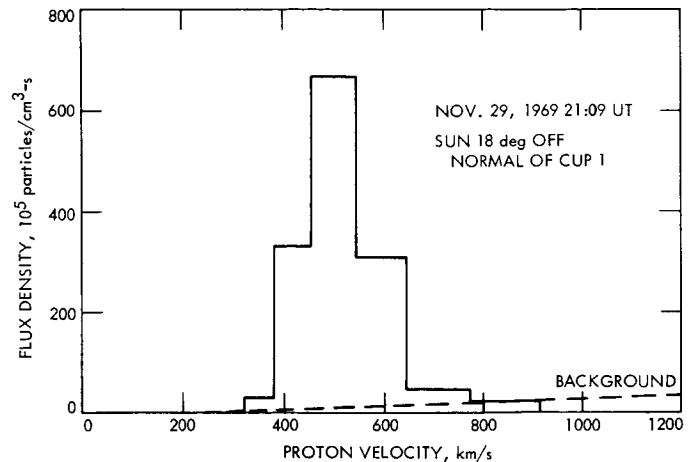
## 7. Types of Spectra Observed

The data discussed in this article have been obtained from the high-speed printer in the ALSEP control room at the Manned Spacecraft Center, and consist of intermittent samples of data from 1 to 8 min in length. The positive-ion spectra observed have been of the general types discussed below.

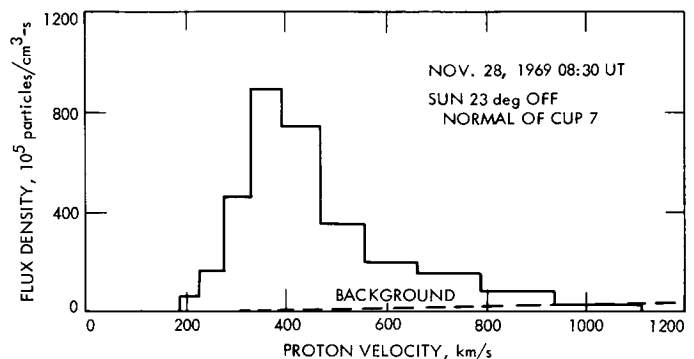
In the first type of spectrum, the peak current is in an energy window corresponding to a velocity of 400 to 550 km/s for protons. The bulk of the remaining currents are in the two adjacent energy windows and a small (4 to 10%) current in the second higher energy window. Figure 8 is a histogram of a typical spectrum of this type. (The background reading of each cup is slightly dependent upon modulation voltage, due primarily to pickup in the electrometer inputs. Until all the data become available and are analyzed, this background is only estimated, giving an uncertainty in readings in the higher velocity windows. The magnitude of this source of uncertainty is indicated in each spectrum as a dashed line marked "background.") This type of spectrum would be expected for the unperturbed (normal interplanetary) solar wind.

The second type of spectrum has a peak current in a window corresponding to a proton velocity of 250 to 450 km/s and significant currents in several adjacent windows on either side of the peak. Such spectra typically have positive ion densities of approximately 5 particles/cm<sup>3</sup>, and adjacent spectra (spaced 28.1 s apart) often show large changes in velocity and density. Double-peaked spectra occasionally appear, probably indicating rapid velocity fluctuations. Figure 9 is a histogram of a typical spectrum included within this type. Such spectra are typical of the transition region between the earth's bow shock and the geomagnetosphere. However, other sources of perturbation, such as solar conditions (at the time the solar wind particles left the sun) or lunar interaction with the solar wind, cannot be ruled out until further information is available.

At times, detectable currents appear in only one energy step or in two adjacent steps so that certain plasma properties (such as velocity distribution) are not calculable for these spectra.



**Fig. 8. Typical positive-ion spectrum obtained in the interplanetary solar wind**



**Fig. 9. Typical positive-ion spectrum obtained in the earth's magnetosheath (perturbed solar plasma)**

During most of the lunar night, there is no detectable flux of solar wind particles within the 100 to 900 km/s range of the instrument.

The electron component of the solar wind has also been detected. However, it is typical of this type of plasma probe that it does not completely distinguish between plasma electrons and photo-electrons so that the interpretation of the data is difficult. Analysis of the complete data from magnetic tape will be required before conclusions can be drawn from the electron data. The times of appearance and disappearance of photo-electrons gave clear indications of sunset and sunrise.

## 8. Observations

At the time of SWS dust cover removal (15:25:30 UT on November 20, 1969) the positive ion spectra observed

were of the second type (perturbed solar wind), and the Lunar Surface Magnetometer (LSM) indicated that the moon was behind the plasma bow shock of the earth.<sup>4</sup> This kind of solar wind data continued with its large fluctuations in bulk velocity and density until about 03:00 UT November 21, when a region of no plasmas was observed. For the next five days, there were only occasional sampling periods of SWS printout, which indicated plasma was present and always of the second type. The LSM indicated that the moon was in the magnetotail.

Commencing about 10:00 UT November 26, 1969, the SWS again entered a region where the large majority of spectra were of the second type (perturbed). From about 12:00 UT on November 28 (when the LSM indicated passage out through the bow shock into interplanetary space) until sunset on December 3, the spectra were mainly of the first type (interplanetary solar wind), although the flux density often became so low as to make classification of spectra type difficult. Several short periods of type-two spectra have been tentatively identified later during November 28.

The SWS was in darkness beginning at 15:22 UT on December 3, as indicated by the cessation of photo-electrons detected in the cup nearest the solar direction. For several hours preceding sunset, the density of the solar wind plasma had appeared to steadily decrease, but this may be partially caused by detector sensitivity reduced at angles near the horizon. The plasma signal continued to decrease until between 19:00 and 20:00 UT on December 3 when the instrument threshold of sensitivity was reached.

In the next 14 days of lunar darkness, no times of plasma detection have been observed from the limited data so far scanned.

The first photo-electrons of sunrise were detected at 11:38 UT on December 18 and within 20 min the first telemetry indication of warming of the sensor assembly was received. However, by sunrise, protons of the solar wind had been observed for nearly 12 h. Several large fluctuations in proton flux occurred, varying from a high flux about 18 min before sunrise to no detectable plasma as late as 15 min after sunrise. These fluctuations may be caused by changes in the solar-wind flux or flow direction or by a lunar interaction effect. Comparison with inter-

planetary solar-wind data will be required to resolve this uncertainty.

Data samples for December 22 to 24 again show no detectable plasma, probably because the instrument was once more in the geomagnetic tail.

## 9. Discussion of Results

This article is based upon the examination of data for the first 35 days after SWS deployment, which is only a few percent of the data that will ultimately be available for this period.

The solar plasma at the lunar surface is superficially indistinguishable from that at a distance from the moon both when the moon is ahead of and behind the plasma bow shock of the earth. No detectable plasma appears to exist in either the earth's magnetotail or in the moon's shadow.

Typically, the preponderance of flux entered only one cup—the one most nearly facing the sun. However, during the time when the SWS was most sensitive to angular variations (from 12:00 UT November 28 to 8:00 UT November 29), there were indications that the direction of bulk velocity varied as much as  $\pm 10$  deg from the mean angle 2 to 4 deg east of the sun's optical direction. This is the only time span for which angular information is now available.

Times of passage through the bow shock or through the magnetotail boundary as indicated by the SWS and by the ALSEP LSM are in agreement at times when comparison of data has been possible.

Observations have generally been in accordance with expectations, but the highly variable spectra observed on November 27 and at sunrise may prove to involve unexpected phenomena.

Complete data and detailed comparison with other solar-wind measurements will be required before firm or quantitative conclusions can be drawn.

## References

1. Colburn, D. S., et al., "Diamagnetic Solar-Wind Cavity Discovered Behind Moon," *Science*, Vol. 158, No. 3804, pp. 1040-1042, Nov. 24, 1967.
2. Sonett, C. P., Colburn, D. S., and Currie, R. G., "The Intrinsic Magnetic Field of the Moon," *J. Geophys. Res.*, Vol. 72, No. 21, pp. 5503-5507, Nov. 1, 1967.

<sup>4</sup>Private communication with P. Dyal of the Ames Research Center.

3. Ness, N. F., et al., "Perturbations of the Interplanetary Magnetic Field by the Lunar Wake," *J. Geophys. Res.*, Vol. 73, No. 11, pp. 3421-3449, June 1, 1968.
4. Ness, N. F., et al., "Early Results from the Magnetic Field Experiment on Lunar Explorer 35," *J. Geophys. Res.*, Vol. 72, No. 23, pp. 5769-5778, Dec. 1, 1967.
5. Lyon, E. F., Bridge, H. S., and Binsack, J. H., "Explorer 35 Plasma Measurements in the Vicinity of the Moon," *J. Geophys. Res.*, Vol. 72, No. 23, pp. 6113-6117, Dec. 1, 1967.
6. Siscoe, G. L., et al., "Experimental Evidence for a Detached Lunar Compression Wave," *J. Geophys. Res.*, Vol. 74, No. 1, pp. 59-69, Jan. 1, 1969.
7. Buehler, F., et al., "Apollo 11 Solar Wind Composition Experiment: First Results," *Science*, Vol. 166, No. 3912, pp. 1502-1503, Dec. 19, 1969.
8. Hundhausen, A. J., "Direct Observations of Solar Wind Particles," *Space Sci. Rev.*, Vol. 8, No. 5/6, pp. 690-749, 1968.

## IV. Communications Systems Research

### TELECOMMUNICATIONS DIVISION

#### A. Coding and Synchronization Studies: Analysis of a Digital Single-Channel Command

System, S. Butman, A. L. Couvillon, R. Goldstein, R. Green, and J. K. Holmes

##### 1. Introduction

The existing *Mariner* command system, which was first successfully employed in 1962, consists of a 1-bit/s PCM/PSK/PM command channel that uses a PN code,<sup>1</sup> transmitted on a separate subcarrier for bit synchronization. The bit sync acquisition time in the *Mariner* Mars 1969 system is about 600 bit-times and is achieved with a reliability of 0.9. Therefore, to ensure that the command has been received and decoded, a verification is relayed to the ground stations. If it has not been decoded, an additional command must be sent. In the Mars missions, this requires an additional 600 s per verification. Obviously, in using this technique on missions to the outer planets, it would take hours whenever a repeat command was required. Moreover, the necessity to retune the DSIF ground transmitter more than once per hour, to compensate for orbital doppler, substantially reduces the time during which commands can be sent to a spacecraft orbiting a planet, as would be the case in the *Mariner* Mars 1971 mission.

A more efficient single-channel analog system has recently been breadboarded and demonstrated at JPL. In this system, bit timing is derived directly from the command bit stream by means of a "bit synchronizer" which can acquire sync in 140 bit-times with a reliability of 0.99. Furthermore, since no power is diverted into a separate sync channel, there is an improvement of about 6 dB in signal-to-noise ratio in the command bit stream, from which it follows that the bit rate can be increased to 4 bits/s so that the acquisition time is reduced to 35 s.

The present single-channel system is all digital, providing both higher reliability than the analog single-channel system and better communication theoretic performance. Analysis shows that it can acquire bit and word sync in a span of 31 bit-times or 7.8 s, with a bit rate of 4 bits/s and a reliability equal to the command bit-error probability of  $1 \times 10^{-5}$ . The significance of this rapid acquisition with high reliability is that the necessity for telemetry verification is eliminated, thereby allowing rapid and reliable command of missions to the outer planets. Moreover, the long duration of such missions (about 12 yr) requires an equipment reliability that is achievable only with digital circuitry.

This article describes the new digital command system, along with the sequence of operations that are required

<sup>1</sup>PCM = pulse-code-modulated, PSK = phase-shift-keyed, PM = phase-modulated, and PN = pseudonoise.

to obtain bit sync, and also develops the associated probabilities of correct operation.

## 2. Description of Digital Command System

This subsection briefly describes the sequence of operations performed by the command receiver to acquire command bit and word synchronization. A flow chart of the operations is given in Fig. 1, and the synchronizing prefix code is depicted in Fig. 2.

To obtain an efficient command channel, the synchronizing prefix code is sent prior to any commands. This provides subcarrier, bit, and word timing to the command receiver for coherent detection of command bits and command words. The synchronizing sequence is designed to ensure correct synchronization of the command receiver with very high reliability since the lack of correct timing results in an unreliable command channel.

Reference to the flow chart (Fig. 1) will show that the command receiver is in reality an adaptive matched filter. At the start, the filter is matched to the subcarrier signal. Once the presence of the subcarrier is detected, its phase is estimated, thus establishing subcarrier timing. The next step is to obtain bit timing. Bit timing is provided by the PN sequence that is sent immediately after the subcarrier signal. The PN sequence is, of course, chopped by the subcarrier and this is why subcarrier timing is necessary prior to detection of the PN code. Reference to the flow chart will again show that the command receiver is now a series of filters matched to the PN code. It, therefore, estimates the phase of the received PN code to provide bit timing. The onset of commands is determined by the receiver when it detects a sign reversal in the PN code, namely,  $\overline{PN}$ . Command bits follow immediately after  $\overline{PN}$ . There are a number of threshold tests and updating operations that are performed to guarantee the necessary reliability. They are discussed in the more detailed description that follows.

As long as the command receiver is turned on, it continually correlates, over intervals of 5 bit-times, each of the 16 equally spaced phases of a locally generated replica of the subcarrier (SC) against the incoming waveform. If the largest correlation exceeds the threshold  $Th_1$ , the receiver decides that a prefix sequence is present, otherwise it is returned to start where it continues to correlate. If the threshold was exceeded, then again all 16 subcarrier phases are correlated for 5 more bit-times. The local SC phase is then set to align itself with the phase corresponding to the largest correlation.

Now that the subcarrier is in sync, the system waits for 10 command bit-times so that the correlators will be correlating against the PN code in order that the receiver can now obtain bit sync. A length-15 PN code (15 cycles of SC) follows the unmodulated portion of the synchronizing code and is used to obtain accurate bit sync. All 15 phases are correlated against the input PN code for 5 bit-times, and the one with the largest correlation is chosen to correspond to the correct bit synchronization. This largest correlation is compared to threshold  $Th_2$ , and if it is exceeded, the local bit timing is aligned to correspond with this largest correlation. And if the threshold is not exceeded, the system is returned to start. This second threshold has the capability of decreasing the probability of a noise-only sequence producing a command and also guards against a transmitter failure on the ground.

Assuming at this point that both the subcarrier and the command bits are in sync, then the next step is to locate the end of the synchronization code. This is accomplished with the use of one bit-time of  $\overline{PN}$ , the complement of PN at the end of the synchronizing code. Since the bit phase is known, only the in-phase correlation need be checked to detect  $\overline{PN}$ . The arrival of  $\overline{PN}$  is signaled by a change in the sign of the correlation from positive to negative. The check could require as many as nine command bit-times or as few as one. The end of the  $\overline{PN}$  component is followed by the start of the first command word.

Finally, the receiver computes the sum of the magnitude of six bit correlations and compares this with threshold  $Th_3$  every six bits. If the sum exceeds the threshold, six more bits are decoded and then again the threshold is checked. If the threshold is not exceeded, the system is returned to start. This mode is needed in order that the receiver will be returned to start when a given sequence of commands from earth has been completed. A new set of commands set at a later time must, of course, be re-synchronized starting back at the beginning of the acquisition sequence.

Assuming the system is properly synchronized, the correlators or matched filters now correlate against the incoming command bits. After 15 cycles of the squarewave subcarrier, the first command bit will then be completed and a decision as to its sign can be made.

During the time from the initial SC synchronization (phase estimate) until the end of  $\overline{PN}$ , there is no up-date of the SC phase estimate. For each command bit following  $\overline{PN}$ , an update or correction is made. This update is either a slip or an advance of  $1/256$  of a SC cycle. This

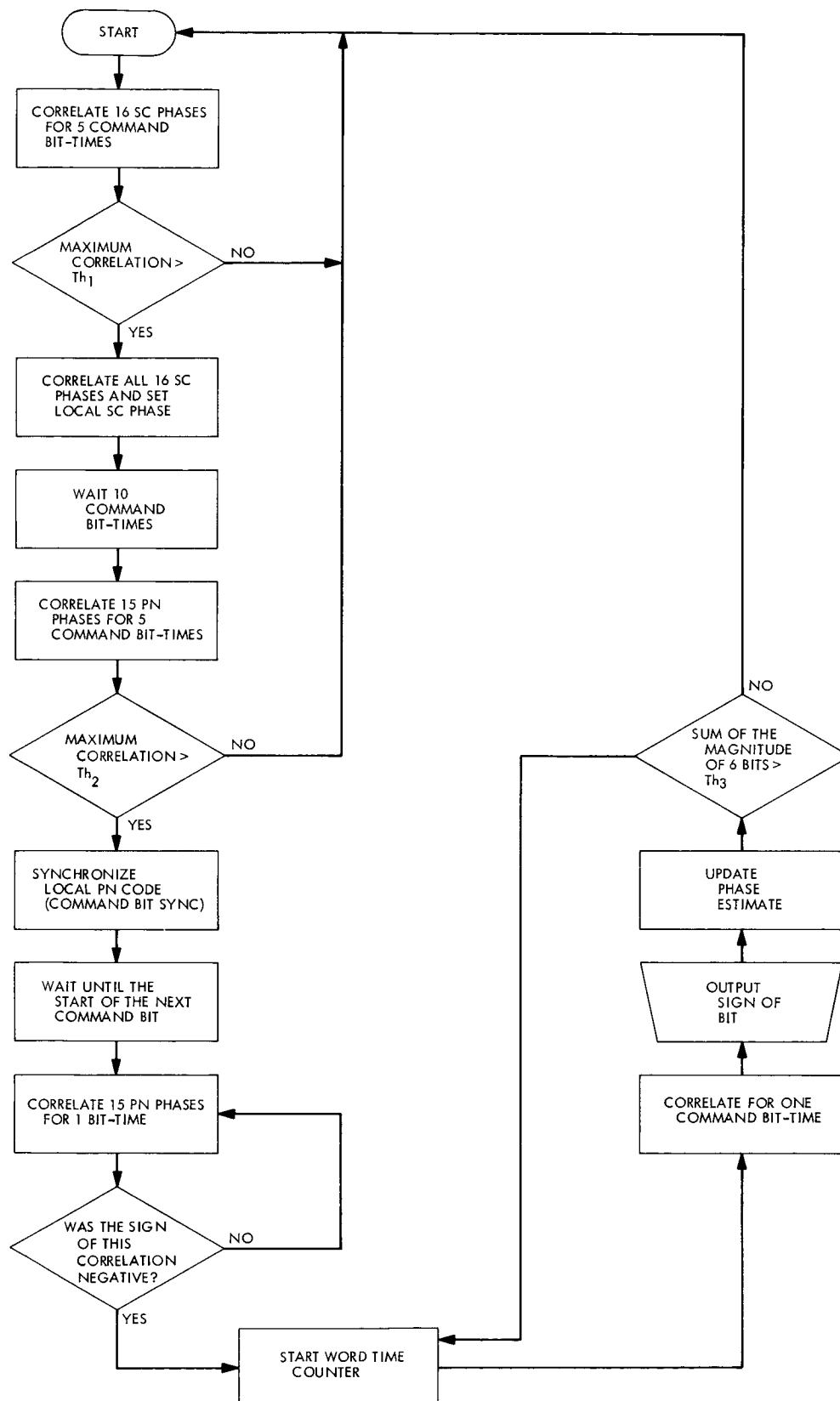
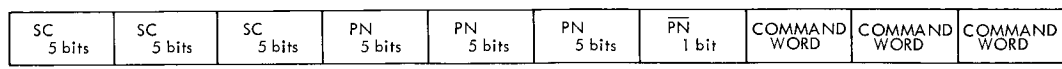
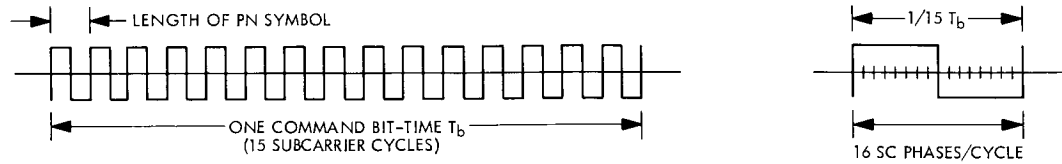


Fig. 1. Flow chart of digital command system

(a) CODE STRUCTURE



(b) SUBCARRIER WAVEFORM AND POSSIBLE PHASE SHIFTS

**Fig. 2. Synchronization code structure and subcarrier waveform**

continual correction insures high reliability for the command word detection process.

### 3. Derivation of Thresholds and Decision Probabilities

In *Subsections 3a-d*, three thresholds are derived so that the associated command system probabilities are compatible with the design requirements. The basic philosophy of design dictated that all anomalies be of probability less than or equal to  $10^{-5}$ . Thresholds, which are required to obtain the correct decision probabilities, have been either analyzed directly or bounded so as to meet the requirement of the probability of an anomaly.

#### *a. Probability of detection and false alarm for the subcarrier.*

**Probability of detection.** The function of the detection mode is to detect the presence of the command prefix code. The detection mode is accomplished by correlating the incoming waveform against 16 equally spaced phase shifts of the squarewave subcarrier. In reality, the system forms the correlations by summing the sampled values. However, to simplify the analysis, it will be assumed that all operations are time-continuous. The difference in performance due to this approximation is negligible. To obtain a bound on the probability of detection  $P_D$ , consider the 16 correlator values  $y_i$ , the  $i$  corresponding to the  $i$ th phase position. It follows that

$$P_D = \text{prob}(y_i \geq Th_1, \text{for some } i) \geq \text{prob}(Y_j \geq Th_1 | \phi_j)$$

where  $\phi_j$  denotes the condition that the incoming waveform is at the  $j$ th phase position. In other words, a lower bound on  $P_D$  is given by the probability that the correlator output, corresponding to the correct phase, is larger than  $Th_1$ , which excludes the possibility that jointly  $Y_j < Th_1$  and  $Y_k \geq Th_1, k \neq j$ , given that the true phase shift is  $\phi_j$ .

If we denote  $s(t - T_j)$  as the signal at time delay  $T_j$  corresponding to the phase shift  $\phi_j$  and denote  $n(t)$  as white gaussian noise with spectral density  $N_0/2$ , then the in-phase correlator value, for  $N$  bit integrations, is given by

$$Y_j = \int_0^{NT_b} [As(t - T_j) + n(t)] s(t - T_j) dt \quad (1)$$

where  $T_b$  is the command bit-time and  $A$  is the signal amplitude with the energy of  $S(t)$  normalized to one. It is easy to show that

$$\text{prob}(Y_j \geq Th_1 | \phi_j) = \Phi \left[ \frac{Th_1 - ANT_b}{\left( \frac{N_0}{2} NT_b \right)^{1/2}} \right] \quad (2)$$

where

$$\Phi(x) \triangleq \int_x^\infty \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{t^2}{2}\right) dt$$

Since a requirement of the design is that

$$P_D \geq 1 - 10^{-5} \quad (3)$$

it follows that we must have

$$\frac{ANT_b - Th_1}{\left( \frac{N_0}{2} NT_b \right)^{1/2}} \geq 4.25$$

**False alarm.** In order that noise alone does not falsely trigger the detection mode, except with very low probability, it is required that the probability of false alarm satisfies

$$P_{FA} \leq 10^{-5}$$

If we denote the correlations with only noise present by  $N_i$ , we have, noting that from the correlation function of a squarewave each noise correlation will have its own negative and, therefore,

$$P_{FA} = P[(|N_1| \geq Th_1) \cup (|N_2| \geq Th_1) \cup \dots \cup (|N_8| \geq Th_1)] \quad (4)$$

where  $A \cup B$  denotes that either  $A$  or  $B$  is true or both are true. By the union bound, we have that

$$P_{FA} \leq \sum_1^{\infty} P(|N_1| \geq Th_1) \quad (5)$$

which can be evaluated as

$$P_{FA} \leq 16 \operatorname{erfc} \left[ \frac{Th_1}{\left( \frac{N_0}{2} NT_b \right)^{1/2}} \right] \quad (6)$$

In order that  $P_{FA} \leq 10^{-5}$ , we must have

$$\frac{Th_1}{\left( \frac{N_0}{2} NT_b \right)^{1/2}} \geq 4.75 \quad (7)$$

Eliminating  $Th_1$  from Eqs. (3) and (7) and using the condition that the command bit-error probability  $PE_b$  be  $10^{-5}$  yields

$$\frac{A^2 T_b}{N_0} = 9, \quad N = 5 \quad (8)$$

Using Eqs. (3), (7), and (8) yields the inequality for  $Th_1$ .

$$2.51 \leq \frac{Th_1}{AT_b} \leq 2.76 \quad (9)$$

All three thresholds will be obtained relative to  $AT_b$ .

**b. PN code detection and its false alarm.** The PN code serves the function of obtaining word sync. A PN code (length 15) is used since, as a class, maximal length PN codes have a normalized correlation of either  $+1$  or  $-1/m$ , where  $m$  is the code word length, and hence are very desirable for synchronization purposes.

**Probability of detection.** We can lower-bound the probability of detection, as in Subsection 3a, by the following:

$$P_D = \operatorname{prob}(Y_i \geq Th_2, \text{ for some } i) \geq \operatorname{prob}(Y_j \geq Th_2 | \phi_j)$$

Or, for  $M$  bits, we then must have, from Eq. (3),

$$\frac{AMT_b - Th_2}{\left( \frac{N_0}{2} MT_b \right)^{1/2}} \geq 4.25 \quad (10)$$

**False alarm.** In this case, only noise is present and so we have

$$P_{FA} = P[(N_1 \geq Th_2) \cup (N_2 \geq Th_2) \cup \dots \cup (N_{15} \geq Th_2)] \quad (11)$$

Again, using the union bound, we have

$$P_{FA} \leq \sum_1^{15} P(N_i \geq Th_2) \quad (12)$$

In order that  $P_{FA} \leq 10^{-5}$ , we must have, after evaluating Eq. (12),

$$\frac{Th_2}{\left( \frac{N_0}{2} MT_b \right)^{1/2}} \geq 4.8 \quad (13)$$

Solving Eqs. (10) and (13) simultaneously and using the fact that  $A^2 T_b / N_0 = 9$  so that  $PE_b = 10^{-5}$  produces

$$2.53 AT_b \leq Th_2 \leq 2.78 AT_b \quad \text{and} \quad M = 5 \quad (14)$$

**c. Detection of  $\overline{PN}$ .** Thus, at this point, subcarrier phase as well as bit sync are presumably acquired, and the remaining step is to locate the start of the actual commands. This is accomplished by detecting the presence of  $\overline{PN}$ , the complement of PN which could occur at any of nine possible times following bit sync. When  $\overline{PN}$  arrives, a negative correlation will appear instead of the positive one associated with PN. The probability of detection is given by

$$P(Y_m \leq 0 | \overline{PN}) \quad (15)$$

where  $Y_m$  is the correlation of the in-phase correlator. The fact that only one command bit-time is needed is clear since the correct detection of a command bit (plus or minus) has the same probability as the correct detection of a PN sequence since both signals have the same energy.

**d. The dump probabilities.** When a command terminates, the system must be returned to start in order that it will be ready to acquire or sync on the next command. Threshold  $Th_3$  is used to decide, upon comparison of a

correlator output, whether the command word has terminated or not.

**Probability of a correct dump.** To decide whether to dump (or return the system to "start"), the sum of the magnitude of  $q$  command bit correlations is added and compared to  $Th_3$ ; if the sum is less than  $Th_3$ , the system is dumped. With the signal absent, we have

$$z = \sum_{i=1}^q |Y_{c_i}| \quad (16)$$

where  $Y_{c_i}$  are the command bit correlations. Without going through the details, it can be shown that the probability of correct dump  $P_{CD}$  satisfies

$$P_{CD} \geq 1 - 2^q \operatorname{erfc} \left[ \frac{Th_3}{\left( \frac{N_0}{2} q T_b \right)^{1/2}} \right] \quad (17)$$

**Probability of false dump.** A false dump is the act of being returned to start by the word detector even though the commands have not terminated. The probability of false dump is given by

$$P_{FD} = P(y < Th_3) \quad (18)$$

where, letting  $q$  denote the number of command bits and  $N_i$  the noise part of the correlation,

$$y = \sum_{i=1}^q |AT_b + N_i| \quad (19)$$

Now since

$$y' = \sum_{i=1}^q (AT_b + N_i) \leq y \quad (20)$$

it follows that

$$P_{FD} \leq P(y' \leq Th_3) \quad (21)$$

It can be shown, evaluating Eq. (21), that

$$\frac{qAT_b - Th_3}{\left( \frac{N_0}{2} q T_b \right)^{1/2}} \geq 4.25 \quad (22)$$

Using Eqs. (17) and (22) with  $A^2 T_b / N_0 = 9$ , we obtain

$$2.92 AT_b \leq \frac{Th_3}{AT_b} \leq 3.55 AT_b, \quad q = 6 \quad (23)$$

So, if  $Th_3$  satisfies Eq. (23), then both the requirements on  $P_{CD}$  and  $P_{FD}$  will be met simultaneously.

#### 4. Overall Command System Probabilities

**a. System probability of accepting a noise-only command.** In order that noise can get through the command system and be interpreted as a legitimate command bit, the following four conditions must be satisfied:

- (1) A noise correlation for the SC phase must exceed  $Th_1$ .
- (2) A noise correlation for the PN phase must exceed  $Th_2$ .
- (3) Noise correlation must be non-positive for the false detection of PN for the correct phase and bit timing.
- (4) The sum of the magnitude of six noise commands must exceed threshold  $Th_3$ .

Since these events depend upon white noise over non-intersecting time intervals, it can be shown, letting  $P_{ANC}$  be the probability of accepting a noise command, that

$$P_{ANC} \leq 10^{-15}$$

and, therefore, is an extremely rare event.

**b. System probability of not detecting the first command bit.** In order for the first command bit not to be detected, at least one of three decisions must be made incorrectly. Therefore, the probability of not detecting the first command  $P_{NDFC}$  is given by

$$P_{NDFC} = P[(\max y^{(1)} \leq Th_1) \cup (\max y^{(2)} \leq Th_2) \cup (\max y^{(3)} \leq Th_3)]$$

where the superscripts denote the correlations associated with the respective thresholds. Since the three events are jointly independent with probability less than or equal to  $10^{-5}$ , the probability of not detecting the first command approximately satisfies

$$P_{NDFC} \leq 3 \times 10^{-5}$$

#### 5. Phase Error Considerations

The probability of the phase error between the incoming and locally generated squarewave subcarrier is presented, assuming that the two squarewaves are either

exactly in phase or a multiple of 1/16 of a cycle apart. When a phase error exists, the signal-to-noise ratio available for command bit detection is decreased, as well as the other functions of the synchronization process which depend on detection of various signals for proper sync.

If we denote the probability of an error in the subcarrier of 1/16 of a cycle by  $PE_{\Delta\phi}$ , then it has been found that

$$PE_{\Delta\phi} \leq 8 \times 10^{-4}$$

where the true phase may be either 1/16 greater or less than the estimated phase. Using similar notation, it is found that the probability of having a phase error of 1/8 of a cycle satisfies

$$PE_{2\Delta\phi} \leq 2.8 \times 10^{-6}$$

and, finally, for an error of 3/16 of a cycle, we have

$$PE_{3\Delta\phi} \leq 9 \times 10^{-10}$$

The effect of the offset between the local SC and the received SC signal is to deteriorate the detected command bit probability. This can be seen by looking at the correlation function for a squarewave. For an offset of 1/16 of a cycle, the correlation function is down to 3/4 of the in-phase value. Hence, during the time the system is locked at a phase offset of 1/16 of a cycle, the signal-to-noise ratio is reduced by a factor of 3/4 or 1.25 dB with the attendant reduction in command bit probability. At a phase offset of 1/8 of a cycle, the signal-to-noise ratio is down to 3 dB from the in-phase value.

## B. Coding and Synchronization Studies: Joint Probability Density of True and Estimated Phase Delay in Sequential Ranging System, S. Butman

### 1. Introduction

To function properly, the new JPL sequential ranging system (SPS 37-52, Vol. II, pp. 46-49) must obtain an accurate estimate of the phase delay in the highest frequency component of the code to provide a clock reference for the remaining, slower, components. Since the components are squarewaves, the problem is one of estimating the phase of a squarewave whose amplitude is unknown and which has been corrupted by additive white

gaussian noise. The formula used for this purpose can be summarized as

$$\hat{\theta} = \frac{\pi}{2} \left[ 1 - \frac{x}{|x| + |y|} \right] \text{sgn } y, \quad -\pi \leq \hat{\theta} \leq \pi \quad (1)$$

where  $\text{sgn } y = 1$  if  $y \geq 0$ ,  $\text{sgn } y = -1$  if  $y < 0$ ,  $|y|$  is the absolute value of  $y$ , and  $x$  and  $y$  are conditionally independent gaussian random variables with conditional means

$$\mathbf{E}[x|\theta] = f(\theta), \quad \mathbf{E}[y|\theta] = f\left(\theta + \frac{\pi}{2}\right)$$

and variances

$$\mathbf{E}[(x - f(\theta))^2 | \theta, \sigma] = \sigma^2 = \mathbf{E}\left[\left(y - f\left(\theta + \frac{\pi}{2}\right)\right)^2 | \theta, \sigma\right]$$

Here  $\theta$  is the true but unknown phase delay and  $f(\theta)$  is the cross correlation function of the received and the locally generated reference signal. Both signals are normalized to unit energy.

The purpose of this article is to derive an expression for the conditional probability density  $p(\hat{\theta}|\theta, \sigma)$  when  $\hat{\theta}$  is computed according to Eq. (1). Since the prior probability density  $p(\theta)$  is uniform by assumption, the joint probability density will be simply  $p(\hat{\theta}, \theta) = p(\hat{\theta}|\theta)/2\pi$ . This result can then be used to analyze and to optimize the performance of the system for various performance measures.

### 2. A Pseudo-Polar Coordinate System

Our results follow in a straightforward way from the observation that Eq. (1) is not so much an estimate as it is a transformation of coordinates. The transformation becomes evident if we note that the pair  $(\hat{\theta}, \hat{z})$  where

$$\hat{z} = |x| + |y|$$

uniquely describes any point in the  $(x, y)$  plane, as illustrated in Fig. 3. We call the  $(\hat{z}, \hat{\theta})$  coordinate system a pseudo-polar system in analogy with the polar coordinates  $(r, \theta)$  where

$$r = (x^2 + y^2)^{1/2}$$

and

$$\theta = \frac{\pi}{2} \left( 1 - \cos^{-1} \left( \frac{x}{r} \right) \right) \text{sgn } y$$

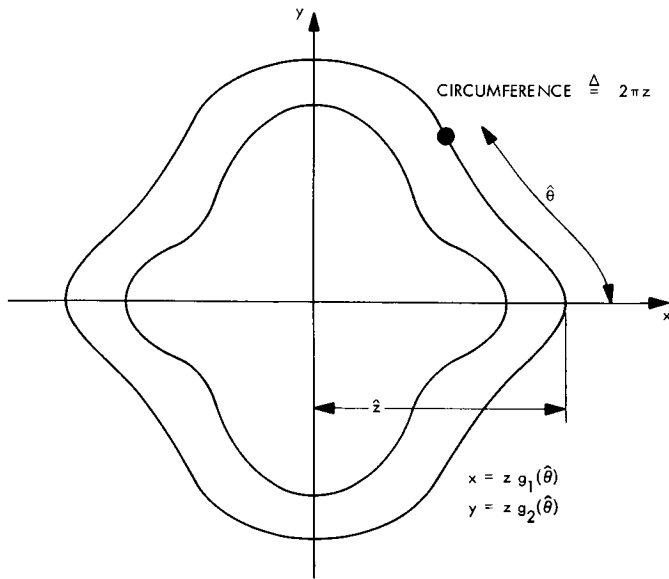


Fig. 3. Pseudo-polar coordinates  $(\hat{z}, \hat{\theta})$  of a point  $(x, y)$  in the plane

By the same analogy, we observe that the parametric equations for polar coordinates

$$x = r \cos \theta$$

$$y = r \sin \theta = r \cos \left( \theta - \frac{\pi}{2} \right)$$

become, in our pseudo-polar system,

$$x = \hat{z} g_1(\hat{\theta}) \equiv \hat{z} g_1(\hat{\theta})$$

$$y = \hat{z} g_2\left(\hat{\theta} - \frac{\pi}{2}\right) \equiv \hat{z} g_2(\hat{\theta})$$

where

$$|g(\theta)| + \left| g\left(\theta - \frac{\pi}{2}\right) \right| = 1 \equiv |g_1| + |g_2|$$

describes a square of diagonal 2 rather than  $\sin^2 \theta + \cos^2 \theta = 1$ , which describes a circle of diameter 2. Moreover, as noted in SPS 37-53, Vol. III, pp. 200-209,  $g(\theta)$  is the autocorrelation of a squarewave of unit amplitude; therefore,  $g(\hat{\theta})$  is a triangular wavefunction specified by

$$g_1(\hat{\theta}) = g(\hat{\theta}) = 1 - \frac{2}{\pi} |\hat{\theta}|, \quad -\pi < \hat{\theta} \leq \pi$$

and

$$g_2(\hat{\theta}) = g\left(\hat{\theta} - \frac{\pi}{2}\right) = [1 - |g_1(\hat{\theta})|] \operatorname{sgn} \hat{\theta}$$

### 3. The Conditional Probability Density $p(\hat{z}, \hat{\theta} | \theta, \sigma)$

We are now in a position to write down the joint conditional density  $p(\hat{z}, \hat{\theta} | \theta, \sigma)$  from the given joint conditional density

$$p(x, y | \theta, \sigma) = \frac{1}{2\pi\sigma^2} \times \exp \left\{ -\frac{1}{2\sigma^2} [(x - f_1(\theta))^2 + (y - f_2(\theta))^2] \right\}$$

Since the transformation from Cartesian to pseudo-polar coordinates is one-to-one, we can transform  $p(x, y | \theta, \sigma)$  into  $p(\hat{z}, \hat{\theta} | \theta, \sigma)$ , as explained in Ref. 1, or any standard probability text. All that is needed is the Jacobian

$$J\left(\frac{x, y}{\hat{z}, \hat{\theta}}\right) \triangleq \det \begin{pmatrix} \frac{\partial x}{\partial \hat{z}} & \frac{\partial x}{\partial \hat{\theta}} \\ \frac{\partial y}{\partial \hat{z}} & \frac{\partial y}{\partial \hat{\theta}} \end{pmatrix} = \frac{\partial x}{\partial \hat{z}} \frac{\partial y}{\partial \hat{\theta}} - \frac{\partial x}{\partial \hat{\theta}} \frac{\partial y}{\partial \hat{z}}$$

Then

$$p(\hat{z}, \hat{\theta} | \theta, \sigma) = p[x(\hat{z}, \hat{\theta}), y(\hat{z}, \hat{\theta})] \left| J\left(\frac{x, y}{\hat{z}, \hat{\theta}}\right) \right|$$

It is easy to verify that for the triangular wavefunction  $g(\hat{\theta})$  given above the absolute value of the Jacobian is  $2\hat{z}/\pi$ , that is,

$$\left| J\left(\frac{x, y}{\hat{z}, \hat{\theta}}\right) \right| = \hat{z} |g_1(\hat{\theta}) g_2'(\hat{\theta}) - g_2(\hat{\theta}) g_1'(\hat{\theta})|$$

$$= \frac{2}{\pi} \hat{z}$$

Consequently,

$$p(\hat{z}, \hat{\theta} | \theta, \sigma) = \frac{1}{\pi^2 \sigma^2} \hat{z} \exp \left\{ -\frac{1}{2\sigma^2} [(\hat{z} g_1(\hat{\theta}) - f_1(\theta))^2 + (\hat{z} g_2(\hat{\theta}) - f_2(\theta))^2] \right\}$$

#### 4. The Condition Density $p(\hat{\theta}|\theta, \sigma)$

To obtain the conditional density  $p(\hat{\theta}|\theta, \sigma)$ , it is necessary only to integrate  $p(\hat{z}, \hat{\theta}|\theta, \sigma)$  over the range of all possible values of  $z$ , from 0 to  $\infty$ . The result, using vector notation, is

$$p(\hat{\theta}|\theta, \sigma) = \frac{2}{\pi^2 \|g\|^2} \exp\left(-\frac{\|f\|^2}{2\sigma^2}\right) \left[ \frac{1}{2} + \left(\frac{\pi}{2\sigma^2}\right)^{1/2} \frac{\langle f, g \rangle}{\|g\|} Q\left(-\frac{\langle f, g \rangle}{(2\sigma^2)^{1/2} \|g\|}\right) \exp\left(\frac{\langle f, g \rangle^2}{2\sigma^2 \|g\|^2}\right) \right]$$

where

$$f = \text{col}(f_1(\theta), f_2(\theta)), \quad g = \text{col}(g_1(\hat{\theta}), g_2(\hat{\theta}))$$

$$\|f\|^2 = f_1^2 + f_2^2, \quad \langle f, g \rangle = f_1 g_1 + f_2 g_2$$

and

$$Q(x) = \frac{1}{(\pi)^{1/2}} \int_x^\infty \exp(-y^2) dy$$

Note that  $p(\hat{\theta}|\theta, \sigma)$  was obtained for any choice of autocorrelation functions  $f(\theta)$  and  $g(\theta)$ . This means that the above expression for  $p(\theta|\theta, \sigma)$  is valid for a statistical description of any class of what may be called pseudopolar phase estimators. Specifically, we note that in the actual ranging system, although  $g(\theta)$  was computed for an ideal squarewave, the received waveform contains only the first and third harmonics. Thus, while

$$g_1(\hat{\theta}) = \frac{8}{\pi^2} \sum_{\substack{k=1 \\ \text{odd}}}^{\infty} \frac{1}{k^2} \cos k\hat{\theta} = 1 - \frac{2}{\pi} |\hat{\theta}|$$

we have

$$f_1(\theta) = \left(\frac{8}{\pi^2} \frac{9}{10}\right)^{1/2} \left[ \cos \theta + \frac{1}{9} \cos 3\theta \right]$$

which is very well approximated by  $1 - 2/\pi |\theta|$ , except at the corners  $\theta = 0, \pm\pi/2$ , and  $\pm\pi$ , the best fit being at  $\theta \approx \pi/4$ . In any case, we can use the expression for  $p(\hat{\theta}|\theta, \sigma)$  to obtain performance data, such as error probability, mean-squared phase error, and phase error bias by means of numerical computations that are now simple to program on a computer.

#### Reference

1. Papoulis, A., *Probability, Random Variables and Stochastic Processes*, Chap. 7. McGraw-Hill Book Co., Inc., New York, 1965.

### C. Coding and Synchronization Studies: On the Behavior of a Phase-Locked Loop During a Temporary Loss of Signal, U. Timor

#### 1. Introduction

In certain coherent communication systems, we are interested in the behavior of the phase-locked loop during a temporary loss of signal. This might be the case whenever skipping cycles by the voltage-controlled oscillator (VCO) will cause an error. Moreover, we might have the same situation without a physical loss of signal. For example, in some ranging systems, we modulate the VCO according to the incoming signal and use this as a loop reference. When the modulation of the received signal is changed, there will be a certain delay until this change has been noticed. During this period, the reference signal in the loop is orthogonal to the received signal, and the control of the VCO will be noise only, as in the case of physical loss of signal.

Let  $\phi$  be the phase between the incoming carrier and the VCO, and  $T$  be the period of effective loss of signal. We will investigate the statistics of  $\phi$  during this period, and find the probability that  $|\phi|$  has not exceeded a given level  $b$ . ( $b = \pi$  corresponds to the case where no cycle has been skipped.)

#### 2. First-Passage-Time Problem

Consider the following problem: Let

$$r(t) = (2)^{1/2} A \cos(\omega t) + n(t)$$

be the input to a first-order phase-locked loop, where  $n(t)$  is white gaussian noise with one-sided spectral density of  $N_0$ . Suppose the signal is removed at time  $t_0$ , where we can assume, without loss of generality, that  $t_0 = 0$ . The probability density of the phase error  $\phi$  at time  $t$ ,  $p(\phi, t|\phi_0)$ , satisfies the differential equation (Ref. 1)

$$\frac{\partial p}{\partial t} = -(\omega - \omega_1) \frac{\partial p}{\partial \phi} + \frac{K^2 N_0}{4} \frac{\partial^2 p}{\partial \phi^2}$$

where  $\omega_1$  is the free frequency of the VCO,  $\phi_0$  ( $|\phi_0| < b$ ) is the phase error at  $t = 0$ , and  $K$  is the loop gain.

We defined the first passage time of level  $\pm b$  as the random variable

$$T = T_b(\phi) = \sup \{t \mid |\phi(\tau)| < b, 0 \leq \tau \leq t\}$$

and denote its probability density by  $f_b(t|\phi_0)$ . Thus, the probability that  $\phi(\tau)$  has not passed the level  $\pm b$  up to time  $T$  is given by

$$P_{\phi_0} = 1 - \int_0^T f_b(t|\phi_0) dt$$

Since  $\phi(t)$  is a continuous Markov process, we can use the following results of Darling and Siegert (Ref. 2):

(1) The Laplace transform of  $p(\phi, t|\phi_0)$  is a product

$$\hat{p}(\phi, \lambda|\phi_0) \triangleq \mathcal{L}[p(\phi, t|\phi_0)] = \begin{cases} u(\phi_0) u_1(\phi), & \phi > \phi_0 \\ v(\phi_0) v_1(\phi), & \phi < \phi_0 \end{cases}$$

In this case,  $\phi(t)$  is a Wiener-Einstein process, the solution of which was derived in Ref. 2:

$$f_b(t|\phi_0) = \frac{D\pi}{b^2} \sum_{n=1}^{\infty} (-1)^{n-1} (2n-1) \cos \left[ \frac{(2n-1)\pi}{2b} \phi_0 \right] \exp \left[ -\frac{D\pi^2 (2n-1)^2}{4b^2} t \right]$$

The probability that  $\phi(t)$  does not pass  $\pm b$  during  $(0, T)$  is given by

$$P_{\phi_0}(T) = 1 - \int_0^T f_b(t|\phi_0) dt = \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)} \cos \left[ \frac{(2n-1)\pi\phi_0}{2b} \right] \exp \left[ -\frac{D\pi^2 (2n-1)^2 T}{4b^2} \right]$$

Let  $b = \pi$ , and denote

$$B = \exp \left( -\frac{DT}{4} \right) = \exp \left( -\frac{K^2 N_0}{16} T \right)$$

Then

$$P_{\phi_0}(T) = \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n-1} \cos \left[ \left( n - \frac{1}{2} \right) \phi_0 \right] B^{(2n-1)^2}$$

and

$$\begin{aligned} \hat{f}_b(\lambda|\phi_0) &\triangleq \mathcal{L}[f_b(t|\phi_0)] \\ &= \frac{v(\phi_0)[u(b) - u(-b)] - u(\phi_0)[v(b) - v(-b)]}{u(b)v(-b) - u(-b)v(b)} \end{aligned}$$

(2)  $u(\phi), v(\phi)$  are any two linearly independent solutions of the differential equation

$$\frac{K^2 N_0}{4} \frac{d^2 y}{d\phi^2} - (\omega - \omega_1) \frac{dy}{d\phi} - \lambda y = 0$$

Thus, to find  $f_b(t|\phi_0)$ , we have to solve this ordinary differential equation and take the inverse transform of  $\hat{f}_b(\lambda|\phi_0)$ .

### 3. Case 1: $\omega = \omega_1$

If the free frequency of the VCO is equal to the input frequency or

$$\frac{(\omega - \omega_1)}{2\pi} T \ll 1$$

$p(\phi, t|\phi_0)$  satisfies the equation

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial \phi^2}, \quad D = \frac{K^2 N_0}{4}$$

(1) If we had a strong signal, that is, if we can assume the initial phase error to be zero ( $\phi_0 \approx 0$ ), we get

$$P(T) = P_0(T)$$

$$\begin{aligned} &= \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n-1} B^{(2n-1)^2} \\ &= \frac{4}{\pi} \left( B - \frac{B^9}{3} + \frac{B^{25}}{5} \cdots \right) \end{aligned}$$

Since  $B < 1$ , the series converges rapidly and in most practical cases we will have to compute very few terms.

*Example.*

$D = 2$  (corresponds to loop signal-to-noise ratio  $\alpha = 12$  and loop bandwidth  $B_L = 6$ )

$$T = 1 \text{ s}$$

$$B = \exp\left(-\frac{DT}{4}\right) = 0.606$$

$$\frac{1}{3} B^3 = 0.0037$$

$$P(T) \approx \frac{4}{\pi} \left( B - \frac{1}{3} B^3 \right) = 0.77$$

In this case,  $\phi(t)$  with probability of 77% will remain within the range  $\pm\pi$  during 1 s after loss of signal.

(2) In general, the initial phase error has a probability density

$$p(\phi_0) = \exp\left(\frac{\alpha \cos \phi_0}{2\pi I_0(\alpha)}\right), \quad |\phi_0| < \pi$$

where  $\alpha$  is the loop signal-to-noise ratio ( $\alpha = A^2/N_0 B_L$ ). To find the probability that  $\phi(t)$  does not exceed  $\pm\pi$ , we

have to average  $P_{\phi_0}(T)$  over  $\phi_0 \in (-\pi, \pi)$ :

$$P(T) = \int_{-\pi}^{\pi} P_{\phi_0}(T) p(\phi_0) d\phi_0 \\ = \frac{4}{\pi^2} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)} B^{(2n-1)^2} A_n(\alpha)$$

where

$$A_n(\alpha) = \frac{1}{I_0(\alpha)} \int_0^{\pi} \cos\left[\left(n - \frac{1}{2}\right)\phi\right] \exp(\alpha \cos \phi) d\phi$$

Since we need only a few terms,  $A_n(\alpha)$  can be easily evaluated.

#### 4. Case 2: $\omega \neq \omega_1$

$p(\phi, t | \phi_0)$  satisfies the differential equation

$$\frac{\partial p}{\partial t} = -(\Delta\omega) \frac{\partial p}{\partial \phi} + D \frac{\partial^2 p}{\partial \phi^2}$$

where  $\Delta\omega = \omega - \omega_1$ . Using the theorems of Darling and Siegert, we get

$$u(\phi) = \exp[(r+s)\phi], \quad v(\phi) = \exp[(r-s)\phi]$$

$$r = \frac{\Delta\omega}{2D}, \quad S = \frac{1}{2D} [(\Delta\omega)^2 + 4D\lambda]^{1/2}$$

and

$$\hat{f}_b(\lambda | \phi_0) = \frac{\exp[r(\phi_0 + b)] \sinh[s(b - \phi_0)] + \exp[r(\phi_0 - b)] \sinh[s(b + \phi_0)]}{\sinh(2sb)}$$

The inversion of this Laplace transform yields

$$f_b(t | \phi_0) = \frac{\pi D}{b^2} \exp\left[-\frac{(\Delta\omega)\phi_0}{2D}\right] \sum_{m=1}^{\infty} m \exp\left\{-\left[\frac{D\pi^2 m^2}{4b^2} + \frac{(\Delta\omega)^2}{4D}\right]t\right\} G_b(m, \phi_0)$$

where

$$G_b(m, \phi_0) = \begin{cases} (-1)^{(m-1)/2} \cosh\left[\frac{(\Delta\omega)b}{2D}\right] \cos\left[\frac{\pi m \phi_0}{2b}\right], & m \text{ odd} \\ (-1)^{(m/2)-1} \sinh\left[\frac{(\Delta\omega)b}{2D}\right] \sin\left[\frac{\pi m \phi_0}{2b}\right], & m \text{ even} \end{cases}$$

The probability that, at time  $T$ ,  $\phi(t)$  has not passed  $\pm b$  is

$$P_{\phi_0}(T) = \frac{\pi D}{b^2} \exp \left[ -\frac{(\Delta\omega)\phi_0}{2D} \right] \sum_{m=1}^{\infty} \frac{m \cdot G_b(m, \phi_0)}{\frac{D\pi^2 m^2}{4b^2} + \frac{(\Delta\omega)^2}{4D}} \exp \left\{ -\left[ \frac{D\pi^2 m^2}{4b^2} + \frac{(\Delta\omega)^2}{4D} \right] T \right\}$$

If  $b = \pi$  and  $B = \exp(-DT/4)$ , we get

$$P_{\phi_0}(T) = \frac{N_0 K^2}{4\pi} \exp \left\{ -\frac{2(\Delta\omega)}{N_0 K^2} \left[ \phi_0 + \frac{(\Delta\omega)T}{2} \right] \right\} \sum_{m=1}^{\infty} \frac{m G_{\pi}(m, \phi_0)}{\frac{(\Delta\omega)^2}{N_0 K^2} + \frac{K^2 N_0 m^2}{16}} B^{m^2}$$

As before, we can average  $P_{\phi_0}(T)$  over  $|\phi_0| < \pi$  and find  $P(T)$ .

If  $\phi_0 \approx 0$ , then

$$G_{\pi}(m, 0) = \begin{cases} (-1)^{(m-1)/2} \cosh \left[ \frac{2\pi(\Delta\omega)}{N_0 k^2} \right], & m \text{ odd} \\ 0, & m \text{ even} \end{cases}$$

and

$$P(T) = \frac{N_0 K^2}{4\pi} \exp \left\{ \left[ -\frac{(\Delta\omega)^2 T}{N_0 k^2} \right] \cosh \left[ \frac{2\pi(\Delta\omega)}{N_0 k^2} \right] \right\} \sum_{n=1}^{\infty} \frac{(-1)^{n-1} (2n-1)}{\frac{(\Delta\omega)^2}{N_0 K^2} + \frac{N_0 K^2 (2n-1)^2}{16}} B^{(2n-1)^2}$$

## 5. Expected Value and Variance of First Passage Time

From  $f_{\pi}(t|\phi_0)$ , we can compute

$$E\{T\} = \int_0^{\infty} t f_{\pi}(t|\phi_0) dt = \frac{N_0 K^2}{4\pi} \exp \left[ -\frac{2(\Delta\omega)\phi_0}{4K^2 N_0} \right] \sum_{m=1}^{\infty} \frac{m G_{\pi}(m, \phi_0)}{\left[ \frac{(\Delta\omega)^2}{N_0 K^2} + \frac{N_0 K^2 m^2}{16} \right]^2}$$

$$E\{T^2\} = \frac{N_0 K^2}{2\pi} \exp \left[ -\frac{2(\Delta\omega)\phi_0}{N_0 K^2} \right] \sum_{m=1}^{\infty} \frac{m G_{\pi}(m, \phi_0)}{\left[ \frac{(\Delta\omega)^2}{N_0 K^2} + \frac{N_0 K^2 m^2}{16} \right]^3}$$

In the case  $\Delta\omega = 0$ ,  $\phi_0 \approx 0$ , this yields

$$E\{T\} = \frac{64}{\pi N_0 K^2} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)^3} = \frac{64}{\pi N_0 K^2} \cdot \frac{\pi^3}{32} = \frac{2\pi^2}{N_0 K^2}$$

$$E\{T^2\} = \frac{8 \cdot 16^2}{\pi (N_0 K^2)^2} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)^5} = \frac{8 \cdot 16^2}{\pi (N_0 K^2)^2} \frac{\pi^5 \cdot 5}{32 \cdot 2 \cdot 24} = \frac{20\pi^4}{3(N_0 K^2)^2}$$

$$\sigma_T^2 = \frac{20\pi^4}{3(N_0 K^2)^2} - \frac{4\pi^4}{(N_0 K^2)^2} = \frac{8}{3} \left( \frac{\pi^2}{N_0 K^2} \right)^2$$

## 6. Conclusion

The behavior of the phase error  $\phi$  during an effective loss of signal of duration  $T$  was expressed in terms of the density function of the first passage time of level  $\pm b$  (where  $b = \pi$  corresponds to the event that at least one cycle has been skipped). Expressions for this density, as well as for the probability that  $\phi$  exceeded  $\pm b$  during  $(0, T)$ , were found in the form of an infinite sum. In practice, the series converges rapidly and only 2 or 3 terms have to be computed.

The results can be useful in the design and evaluation of systems (such as ranging) where we encounter an effective loss of signals and skipping cycles can cause an error.

## References

1. Viterbi, A. J., *Principles of Coherent Communication*. McGraw-Hill Book Co., Inc., New York, N.Y., 1966.
2. Darling, D. A., and Siegert, A. J. F., "The First Passage Problem for a Continuous Markov Process," *Ann. Math. Statist.*, Vol. 24, pp. 624-639, 1953.
3. Jolly, L. B. W., *Summation of Series*. Dover Publications, Inc., New York, N.Y., 1961.

## D. Coding and Synchronization Studies: The Sensitivity of Sequential Decoders to Noise Correlation, J. E. Savage<sup>2</sup>

### 1. Introduction

Sequential decoding is an effective decoding procedure for memoryless channels which are very noisy since it allows operation at data rates which are substantial fractions of channel capacity. Equivalently, it permits signaling at signal-to-noise ratios which are near the minimum required by channel capacity.

The radio channel created by a spacecraft and a receiving station is essentially memoryless, and existing theory and experimentation covering the frequency of buffer overflow and the undetected error rate for sequential decoding (Refs. 1-3) apply directly to this channel. However, if a ground link between a receiving station and a sequential decoder is used, noise introduced by this link will generally be correlated and modifications in theory are necessary. The purpose of this article is to develop a heuristic argument which shows that memory will cause

an increase in the buffer overflow probability or an effective reduction in computed rate  $R_{\text{comp}}$ . We also present a formula which can be used to calculate  $R_{\text{comp}}$  on channels with memory.

### 2. Heuristic Argument

A tree code of rate  $R = (\log_2 b)/n_1$  (bits per symbols) has  $b$  branches at its base and  $n_1$  symbols per branch, and each of these branches terminates in a node from which  $b$  additional branches originate, etc. The tree is said to have length  $L$  if every full length path through the tree contains  $L$  branches. The number of paths of length  $r$  in the tree is  $b^r$ ,  $1 \leq r \leq L$ .

A tree path  $u$  is transmitted, sequence  $v$  is received, and the sequential decoder searches a copy of the tree code at the decoder to find a path  $u^*$  which is most likely to have been sent, given that  $v$  has been received. This search is simplified if the code is a convolutional code which can be generated with a few shift registers. The probability that  $u$  and  $u^*$  disagree can be made very small.

Sequential decoders are limited by the amount of searching that has to be done to find  $u^*$ . The number of decoding computations is a random variable which has a Pareto distribution. For the discrete memoryless channel (DMC), this implies that the probability of overflowing a buffer which stores  $B$  branches behaves as

$$P_{BF}(L) = L \frac{1}{(B\mu)^{\alpha(R)}}$$

where  $\mu$  is the speed-factor or the maximum number of decoder computations per received symbol and  $L$  is the number of decoded branches.  $\alpha(R)$  is the Pareto exponent, which is a function of the code rate  $R$  and channel parameters. The Pareto exponent is a dimensionless quantity which has value 1 when  $R = R_{\text{comp}}$  and larger values when  $R < R_{\text{comp}}$ . Rates above  $R_{\text{comp}}$  are avoided because a doubling of the buffer size or decoder speed has less than a doubling effect on  $P_{BF}(L)$ .

The probability  $P_{BF}(L)$  is proportional to  $L$  because there are  $L$  opportunities for the decoder to start a long search requiring  $B\mu$  or more computations, which results in buffer overflow. Consider the number of computations  $C_i$  required to decide which branch at node  $i$  is the correct branch, assuming no prior decoding errors. Assume also, for the sake of argument, that the channel is binary and that the received sequence  $v_r$  of  $r$  branches can be written as  $v_r = u_r + e_r$ , where  $e_r$  is the error sequence

<sup>2</sup>Consultant from the Division of Engineering, Brown University, Providence, Rhode Island.

of  $r$  branches ( $n_1 r$  symbols) and  $+$  indicates term-by-term modulo 2 addition of digits.

Assume that the density of errors is uniform over the  $n_1 r$  received symbols and equal to  $d$ . If  $d$  is larger than half of the minimum distance of the code, the decoder will essentially have to examine all branches on paths of length  $r$  or less, or  $(b^{r+1} - 1)/(b - 1)$  branches, before making a permanent decision on a branch at node  $i$ . Therefore, in this case,

$$C_i(e) \cong \frac{b}{b-1} b^r = 2^{n_1 r R} \quad (1)$$

which grows exponentially with the number of received symbols  $n_1 r$ . The number of errors  $d$  will be proportional to  $n_1 r$ , say, or

$$d = \lambda n_1 r \quad (2)$$

If the channel is binary and symmetric with crossover probability  $p$ , we have

$$\begin{aligned} P(e) &\cong \left( \frac{n_1 r}{\lambda n_1 r} \right) p^{\lambda n_1 r} q^{(1-\lambda) n_1 r} \\ &\cong (2^{H(\lambda)}) p^{\lambda} q^{1-\lambda} n_1 r \end{aligned} \quad (3)$$

which decreases exponentially if  $\lambda > p$ , or if  $d$  is greater than the average number of received errors. Now substituting for  $n_1 r$  from Eq. (1) into Eq. (3), we have

tuting for  $n_1 r$  from Eq. (1) into Eq. (3), we have

$$P(C_i) \cong C_i^{-\alpha} \quad (4)$$

where

$$\alpha = \frac{-[H(\lambda) + \lambda \log_2 p + (1-\lambda) \log_2 q]}{R} \quad (5)$$

which is the Pareto distribution. The Pareto exponent is only an approximation to the correct exponent.

Should the channel have memory which causes error bursts to occur, the probability  $P(e)$  will decrease more slowly with  $d$  for large  $d$  than Eq. (3) does, and the result is that  $P(C_i)$  decreases slowly with  $C_i$  for large values of  $C_i$ . Consequently, buffer overflows occur more frequently and the effective  $R_{\text{comp}}$  will be reduced.

Again, this phenomenon occurs because the number of tree branches that must be searched grows exponentially with the length of a noisy channel interval.

### 3. $R_{\text{comp}}$ for Channels with Memory

In Ref. 4, the ensemble average number of computations to decode a branch with the Fano algorithm (Ref. 5) is bounded. From Eqs. (3.15) and (3.19) of Ref. 4, we have

$$C_i \leq (b+1) \frac{2^{T_0}}{2^{T_0/2} - 1} \left\{ \sum_{r=0}^L 2^{r n_1/2} [U - R_{\text{comp}}(r n_1)] \right\} \left\{ \sum_{s=0}^L 2^{s n_1} [2R - U - R_{\text{comp}}(s n_1)] \right\} \quad (6)$$

where  $T_0$  is the separation between thresholds and  $U$  is the bias in the path metric given by Eq. (3.1). Also,  $R_{\text{comp}}(n)$  is given by

$$R_{\text{comp}}(n) = \max_{\{P(u_n)\}} \left( -\frac{1}{n} \log_2 \sum_{v_n} \left\{ \sum_{u_n} P(u_n) \left[ P\left(\frac{v_n}{u_n}\right) \right]^{1/2} \right\}^2 \right) \quad (7)$$

where  $u_n$  and  $v_n$  are  $n$ -symbol input and output sequences, respectively,  $\{P(u_n)\}$  is a probability assignment to input sequences, and  $\{P(v_n|u_n)\}$  is the set of transition probabilities between input and output sequences which is defined by the channel model. For the DMC, Eq. (7) gives the standard expression for  $R_{\text{comp}}$ .

If  $R_{\text{comp}}(n)$  approaches a constant  $R_{\text{comp}}^0$  for large  $n$ , convergence of Eq. (6) for large  $L$  is guaranteed when

$$U < R_{\text{comp}}^0 R < \frac{U}{2} + \frac{R_{\text{comp}}^0}{2} \quad (8)$$

or when

$$R < R_{\text{comp}}^0 \quad (9)$$

Expression (7) has been evaluated in Ref. 5 for a binary channel in which the noise is additive and produced by the two-state Markov source shown in Fig. 4. A channel error is introduced only when the source enters the "1" state. Choosing  $P_1$  and  $P_2$  to produce error clustering, it is found that  $R_{\text{comp}}$  decreases, for small clustering, from the value that it would have for a binary symmetric channel with equal average error rate.

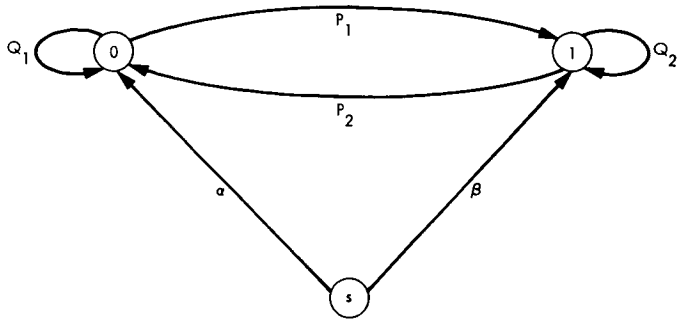


Fig. 4. Markov noise model

Results for this noise source are of limited value for large clustering since the state sequence of the source can then be tracked with high probability and essentially error-free coding would be possible. Nevertheless, Eq. (7) is still useful since it can be applied to a variety of channel models and used with experimental data.

#### References

1. Savage, J. E., "Sequential Decoding—The Computation Problem," *Bell System Technical Journal*, Vol. 45, No. 1, pp. 149–175, 1966.
2. Jordan, K. L., Jr., "The Performance of Sequential Decoding in Conjunction with Efficient Modulation," *IEEE Trans.*, Vol. COM-14, pp. 283–296, 1966.
3. Jacobs, I. M., and Berlekamp, E. R., "A Lower Bound to the Distribution of Computation for Sequential Decoding," *IEEE Trans.*, Vol. IT-13, pp. 167–174, 1967.
4. Savage, J. E., "Progress in Sequential Decoding," in *Advances in Communication Systems*, Vol. 3. Edited by A. V. Balakrishnan. Academic Press, New York, N.Y., 1968.
5. Savage, J. E., *Sequential Decoding on a Channel with Simple Memory*, Group Report. Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., 1964.

### E. Coding and Synchronization Studies: On the Choice of Frequencies for Squarewave Subcarriers, C. L. Weber<sup>3</sup>

#### 1. Introduction

When coherent digital systems employ multiple square-wave subcarriers, such as in a multiple-access satellite, the relative choice of squarewave fundamental frequencies should be made to minimize several quantities (simultaneously, if possible). The distortion power, i.e., that

power which is unrecoverable due to the phase modulation process, should be minimized; the power in the data bearing subcarriers which is spread into the region of the carrier tracking loop should be minimized, or, if possible, eliminated. Finally, the power from each subcarrier which is in the frequency band of the other subcarriers, causing interchannel interference, should be minimized, or, if possible, eliminated.

It is herein shown that all three of these interfering or unrecoverable powers can be simultaneously minimized or eliminated for the case of two squarewave subcarriers. This is accomplished by choosing the second subcarrier frequency to be an even multiple of the first subcarrier frequency.

#### 2. Squarewave Subcarrier Model

In the transmitter, where two subcarrier channels are assumed, the data signals  $\{s_k(t), k = 1, 2\}$  biphasemodulate the frequency-multiplexed squarewave subcarrier waveforms

$$\text{sq}(\omega_k t) \triangleq \text{sgn}(\sin \omega_k t), \quad k = 1, 2$$

Thus,  $\text{sq}(\omega_k t)$  is a squarewave with fundamental frequency  $f_k = \omega_k/2\pi$  and unity amplitude. The input to the carrier phase modulator is given by

$$\theta(t) = \sum_{k=1}^2 m_k s_k(t) \text{sq}(\omega_k t)$$

where  $m_k$  is the modulation index for the  $k$ th subcarrier waveform before carrier phase modulation.

Inherent in the model is the fact that the subcarriers are 100% modulated. Therefore, the subcarrier phase and bit sync are obtained directly from the data signal.

The data signals  $\{s_k(t)\}$  consist of a sequence of  $\pm 1$ 's with bit times  $T_k$ ,  $k = 1, 2$ , respectively. The rate in each subcarrier channel is assumed to be much less than the clock rate in the subcarrier squarewaves, i.e.,

$$\frac{1}{T_k} < \frac{\omega_k}{2\pi}, \quad k = 1, 2 \quad (1)$$

The transmitted signal is given by

$$s(t) = (2P)^{1/2} \sin(\omega_0 t + \theta(t) + \theta_0)$$

<sup>3</sup>Consultant, Electrical Engineering Department, University of Southern California, Los Angeles, California.

where  $P$  is the overall average transmitted power and  $\theta_0$  is some unknown constant reference phase.

The received waveform is

$$y(t) = s(t) + n(t)$$

where  $n(t)$  is assumed white gaussian noise with one-sided spectral density  $N_0$  in watts/Hertz.

The coherent carrier tracking loop generates the reference signal

$$r(t) = 2^{1/2} \cos [\omega_c t + \hat{\theta}_0(t)]$$

Therefore, the output data bearing waveform  $y_0(t)$  from the carrier tracking loop, which is the input to the sub-carrier demodulators, is given by

$$y_0(t) = s_0(t) + n_0(t)$$

where

$$s_0(t) = P^{1/2} \sin [\theta(t) + \phi_r(t)]$$

in which

$$\phi_r(t) \triangleq \theta_0 - \hat{\theta}_0(t)$$

is the carrier-loop phase error, and the additive noise  $n_0(t)$  has the same statistics as  $n(t)$ .

### 3. Carrier Power

With no detuning assumed, the steady-state mod  $2\pi$  probability density function of  $\phi_r$  is given approximately<sup>4</sup> by

$$p(\phi_r) \approx \frac{\exp(\alpha_r \cos \phi_r)}{2\pi I_0(\alpha_r)}, \quad -\pi \leq \phi_r \leq \pi$$

where

$$\alpha_r = \frac{P_c}{N_0 B_L} \quad (2)$$

is the signal-to-noise ratio of the carrier tracking loop, and  $B_L$  is the one-sided noise bandwidth of the carrier tracking loop based on the linear theory. The average power of the received signal at the carrier frequency  $P_c$  is given by the average power of

$$(2P)^{1/2} \sin [\omega_c t + m_1 \text{sq}(\omega_1 t + \theta_1) + m_2 \text{sq}(\omega_2 t + \theta_2) + \theta_0]$$

at the carrier frequency  $\omega_c$ . On account of Eq. (1), the spreading effects of the data signals are neglected.

The expression for  $P_c$  can be written as

$$\left(\frac{P_c}{P}\right)^{1/2} = \cos m_1 \cos m_2 - \sin m_1 \sin m_2 \left( \left\{ \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{4}{k\pi} \sin [k\omega_1(t - t_1)] \right\} \left\{ \sum_{\substack{l=1 \\ l \text{ odd}}}^{\infty} \frac{4}{l\pi} \sin [l\omega_2(t - t_2)] \right\} \right) \Big|_{\text{dc}} \quad (3)$$

where the notation  $|_{\text{dc}}$  means the dc component of, and the  $\{t_i, i = 1, 2\}$  are arbitrary time references. Expressing the double sum in terms of the sums and differences of cosine terms, we find that the sum terms have no dc component and disappear, so that the term enclosed by long parentheses in Eq. (3) can be expressed as

$$\frac{8}{\pi^2} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \sum_{\substack{l=1 \\ l \text{ odd}}}^{\infty} \left[ \frac{1}{kl} \cos (k\omega_1 - l\omega_2) t - (k\omega_1 t_1 + l\omega_2 t_2) \right] \quad (4)$$

The constant term in this sum represents the loss in power into the carrier tracking loop due to interchannel interference between the two squarewave subcarriers. This dc component can be made to vanish if  $(k\omega_1 - l\omega_2)$  is never zero. Since  $k$  and  $l$  are always odd integers, if  $\omega_2$  is an even multiple of  $\omega_1$ , there will be no dc component. By

setting  $\omega_2 = K\omega_1$ , where  $K$  is an even positive integer,

$$(k\omega_1 - l\omega_2) = (k - lK)\omega_1$$

never vanishes.

<sup>4</sup>A second-order loop is assumed.

Assume, henceforth, the subcarrier frequencies are so chosen that

$$\left(\frac{P_c}{P}\right)^{1/2} = \cos m_1 \cos m_2 \quad (5)$$

and

$$\alpha_r = \frac{P_c}{N_0 B_L} = \beta_r \cos^2 m_1 \cos^2 m_2 \quad (6)$$

where

$$s_0(t) = P^{1/2} \sin [m_1 s_1(t) \text{sq}(\omega_1 t + \theta_1) + m_2 s_2(t) \text{sq}(\omega_2 t + \theta_2) + \phi_r(t)] \quad (8)$$

Each subcarrier tracking loop generates an estimate of the phase of the incoming subcarrier clock for its channel, which is an input to the phase detector as shown in Fig. 5.

Considering subcarrier channel 1, the subcarrier phase detector output  $\epsilon_1(t)$  is

$$\epsilon_1(t) = P^{1/2} \text{sq}\left(\omega_1 t + \frac{\pi}{2} + \hat{\theta}_1\right) \sin [m_1 s_1(t) \text{sq}(\omega_1 t + \theta_1) + m_2 s_2(t) \text{sq}(\omega_2 t + \theta_2) + \phi_r(t)] + \text{noise} \quad (9)$$

Expanding, the signal portion of Eq. (9) reduces to

$$P^{1/2} \left[ \cos \phi_r \cos m_2 \sin m_1 s_1(t) \text{sq}(\omega_1 t + \theta_1) \text{sq}\left(\omega_1 t + \frac{\pi}{2} + \hat{\theta}_1\right) + \cos \phi_r \cos m_1 \sin m_2 s_2(t) \text{sq}(\omega_2 t + \theta_2) \text{sq}\left(\omega_1 t + \frac{\pi}{2} + \hat{\theta}_1\right) \right. \\ \left. + \sin \phi_r \cos m_1 \cos m_2 \text{sq}\left(\omega_1 t + \frac{\pi}{2} + \hat{\theta}_1\right) - \sin \phi_r \sin m_1 \sin m_2 s_1(t) s_2(t) \text{sq}(\omega_2 t + \theta_2) \text{sq}\left(\omega_1 t + \frac{\pi}{2} + \hat{\theta}_1\right) \right] \quad (10)$$

The desired signal is the first term in Eq. (10). Under the assumption the subcarrier tracking loop is adequately locked to the signal  $\text{sq}(\omega_1 t + \theta_1)$ , the desired term is

$$P^{1/2} \cos \phi_r \cos m_2 \sin m_1 s_1(t) \left[ 1 - \frac{2}{\pi} |\phi_1| \right]$$

where  $\phi_1 \triangleq \theta_1 - \hat{\theta}_1$ . The second and fourth terms in

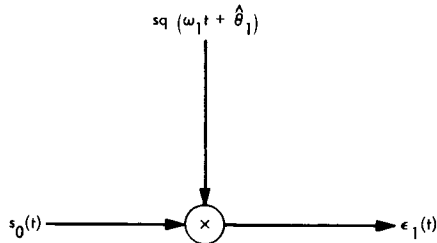


Fig. 5. Subcarrier phase detector (channel 1)

where

$$\beta_r \triangleq \frac{P}{N_0 B_L} \quad (7)$$

is the overall input signal-to-noise ratio.

#### 4. Interchannel Interference

The signal into the subcarrier tracking loops is

$$y_0(t) = s_0(t) + n_0(t)$$

Eq. (10) will not have any components in the low-pass frequency range, where the first term is located, if  $\omega_1$  and  $\omega_2$  are related by an even integer. The third term will always be outside the frequency range of the desired signal.

Therefore, all interchannel interference terms will not be located in the frequency range of the desired signal if  $\omega_2 = K\omega_1$ , where  $K$  is an even integer.

#### 5. Error Rate

Under the assumption the subcarrier tracking loop is performing perfectly, the signal into the first demodulator is given by

$$y_1(t) = P^{1/2} \cos \phi_r \cos m_2 \sin m_1 s_1(t) + n_1(t) \quad (11)$$

where  $n_1(t)$  has the same statistics as  $n(t)$ . The probability of error in the first subcarrier channel is approximately given by

$$P_e = \operatorname{erfc} \left( \frac{2\mathcal{E}_1}{N_0} \right)^{1/2} \quad (12)$$

where

$$\frac{\mathcal{E}_1}{N_0} = \frac{\beta_r}{\delta_1} \cos^2 m_2 \sin^2 m_1 [E(\cos \phi_r)]^2 \quad (13)$$

in which

$$\delta_1 = \frac{1}{B_L T_1} \quad (14)$$

## F. Coding and Synchronization Studies: Signal Design for Fast-Fading Channels,

A. V. Balakrishnan<sup>5</sup>

### 1. Introduction

In the near future, with projects such as *Viking*, the problems of communicating in fast-fading channels will become a reality. To communicate effectively in this kind of channel, its characteristics will have to be taken into account. This article continues to study the signal design problem for fast-fading channels that was initiated in Ref. 1. In fast-fading channels, the received signal contains, in addition to the usual ambient-temperature-limited (white gaussian) noise, a fast-fading or clutter component. This component may be expressed as the output of a random time-varying linear filter with the transmitted signal as the input. For a given transmitted signal, there is always an optimum receiver that maximizes the probability of detection whether in the radar context or the pulse-code-modulated (PCM) communications context. The signal design problem is that of choosing the best signal waveform that provides the maximum probability of detection. Such a problem was considered in the radar or sonar context in Ref. 1 for the case of stationary fast-fading.

In this article, we consider the most general fast-fading channel and show that the receiver need not be optimized for the transmitted signal, and that the optimal overall system can be chosen from the class of systems using

matched filter receivers. We show that if there is no bandwidth constraint on the signals to be used, the problem is actually degenerate in the sense that it is always possible to find a sequence of signals occupying increasing bandwidth and in the limit obtain the best possible performance, the same as if there was no fast-fading component present. For the case of stationary fast-fading, we show that if the signals are restricted to a fixed frequency band, then there is an optimal signal using a matched filter receiver which has the property that the corresponding fast-fading frequency spectrum is uniform (flat) across the band. We also indicate how the results apply to the problem of coherent biphasic PCM transmission of signals from a spacecraft on the vicinity of a planet, the fast-fading component now arising from surface scattering by the planet.

### 2. Main Results (General Case)

We present the main mathematical results for the general (not necessarily stationary fast-fading) case. Let  $s(t)$  denote the transmitted signal, assumed to be narrow-banded about some center carrier frequency  $f_c$ . We shall find it convenient to use the complex (or "analytic signal") notation, so that we can write

$$s(t) = \begin{cases} \operatorname{Re} S(t) \exp(2\pi i f_c t), & 0 < t < T \\ 0, & \text{otherwise} \end{cases}$$

Here  $T$  denotes the duration of the signal, assumed finite. It is also convenient to introduce the real Hilbert space  $\mathcal{H}$  of complex-valued functions  $F(\cdot)$  such that

$$\int_0^T |F(t)|^2 dt < \infty$$

with the inner product defined by

$$[F, G] = \operatorname{Re} \int_0^T F(t) \overline{G(t)} dt$$

where the overbar denotes the complex conjugate. We assume that  $S(\cdot)$  is in  $\mathcal{H}$ , and that

$$[S, S] \leq E$$

The fast-fading return noise component can be expressed as

$$n_c(t) = \operatorname{Re} N_c(t) \exp(2\pi i f_c t)$$

<sup>5</sup>Consultant, Engineering Department, UCLA, Los Angeles, California.

where

$$N_c(t) = \int_0^\infty K(t, \sigma) S(t - \sigma) d\sigma = \int_0^t K(t, t - \sigma) S(\sigma) d\sigma \quad (1)$$

where  $K(t, \sigma)$  denotes the complex impulse response of the time-varying linear system corresponding to the fast-fading noise,  $K(t, \sigma)$  being a random function of the variables, and defined to be zero for  $\sigma < 0$ . Here  $N_c(t)$  is then the "complex envelope" of the fast-fading noise. Similarly, let  $N(t)$  denote the complex envelope of the limiting thermal noise. Then the complex envelope of the received signal can be expressed as

$$X(t) = as(t - L) + N_c(t) + N(t), \quad t > L \quad (2)$$

where  $a$  (in the usual set-up) denotes the unknown signal strength and  $L$  is the propagation delay. The action of any "linear" receiver can be described in terms of the

test statistic:

$$\operatorname{Re} \int_L^{L+T} X(t) \overline{W(t - L)} dt$$

where  $W(\cdot)$  is an element of  $\mathcal{H}$ . The corresponding signal output is

$$a[S, W]$$

and the noise output variance is the variance of

$$\operatorname{Re} \int_L^{L+T} (N_c(t) + N(t)) \overline{W(t - L)} dt$$

Assuming the function  $K(t, \sigma)$  to have zero mean, this variance can be expressed as

$$[R(S)W, W]$$

where  $R(S)$  is the linear-bounded operator mapping  $\mathcal{H}$  into  $\mathcal{H}$  defined by

$$R(S)G = H, \quad H(t) = \frac{1}{2} \left( \int_0^T R_1(t, s) G(s) ds + \int_0^T R_2(t, s) \overline{G(s)} ds \right)$$

where

$$R_1(t, s) = E \left( \overline{(N_0(t + L) + N(t + L))} (N_c(s + L) + N(s + L)) \right)$$

$$R_2(t, s) = E \left( (N_0(t + L) + N(t + L)) \overline{(N_c(s + L) + N(s + L))} \right)$$

and  $E$  denotes the expectation operator. Now, from Eq. (1),

$$E(\overline{N_c(t)} N_c(s)) = \int_0^\infty \int_0^\infty E \left( \overline{K(t, \sigma_1)} K(s, \sigma_2) \right) \overline{S(t - \sigma_1)} S(s - \sigma_2) d\sigma_1 d\sigma_2 \quad (3)$$

$$E(N_c(t) N_c(s)) = \int_0^\infty \int_0^\infty E \left( K(t, \sigma_1) K(s, \sigma_2) \right) S(t - \sigma_1) S(s - \sigma_2) d\sigma_1 d\sigma_2 \quad (4)$$

and we are assuming that the fast-fading noise is independent of the thermal noise, the latter being white gaussian with spectral density  $\phi$ , so that we can finally write

$$[R(S)W, W] = [R_c(S)W, W] + \phi[W, W]$$

where  $R_c(S)$  is the fast-fading covariance operator defined by

$$R_c(S)W = H, \quad H(t) = \frac{1}{2} \left( \int_0^T r_1(t, s) W(s) ds + \int_0^T r_2(t, s) \overline{W(s)} ds \right)$$

where

$$r_1(t, s) = E(\overline{N_c(t+L)} N_c(s+L))$$

$$r_2(t, s) = E(N_c(t+L) N_c(s+L))$$

The signal-to-noise ratio can then be expressed as

$$\frac{a^2 [S, W]^2}{[R_c(S) W, W] + \phi [W, W]} \quad (5)$$

As is well known (Ref. 2), the probability of detection is a monotone-increasing function of this ratio if the fast-fading component can be assumed to be gaussian. For fixed  $S(\cdot)$ , we may thus maximize the signal-to-noise ratio to obtain the optimal receiver. Having chosen the optimal receiver, we may then choose the optimal signal subject to the energy constraint, to maximize the signal-to-noise ratio. This is the signal design philosophy considered in Ref. 1. Alternately, it can also be described as the variational problem:

$$\sup_{[S, S] \leq E} \sup_{W \in \mathcal{H}} \frac{a^2 [S, W]^2}{[R_c(S) W, W] + \phi [W, W]} \quad (6)$$

Since for each  $S(\cdot)$ ,  $R_c(S)$  is non-negative definite, we always have that an upper bound to the above signal-to-noise ratio is

$$\frac{a^2 E}{\phi} \quad (7)$$

If, for some non-zero  $S(\cdot)$  in  $\mathcal{H}$ , we have (the "trivially singular" case<sup>6</sup>)

$$[R_c(S) S, S] = 0 \Leftrightarrow R_c(S) S = 0 \quad (8)$$

then Expression (7) is achieved, and only then. Our main result is that we can always find a sequence of elements  $S_n(\cdot)$  in  $\mathcal{H}$  with energy  $E$ , such that the corresponding signal-to-noise ratio can be made as close to Expression (7) as desired, using only "matched-filter" receivers. For a given signal  $S(\cdot)$ , there is always an optimal receiver  $W(\cdot)$  that maximizes Expression (5); indeed, the optimal  $W(\cdot)$  is given by

$$(R_c(S) + \phi I)^{-1} S$$

<sup>6</sup>The terminology used in Ref. 3.

as is well known (Ref. 2). But this depends on the fast-fading covariance which may not be known. A "matched-filter" receiver is one in which we take

$$W = S$$

which is thus suboptimum, in general, but has the desirable feature that no knowledge of the fast-fading characteristics is required.

**THEOREM 1.** Assume that

$$E(|K(t, s)|^2) \quad (9)$$

is locally square-integrable. Then we can always find a sequence of elements  $S_n$  in  $\mathcal{H}$  such that

$$[S_n, S_n] = E$$

and such that the signal-to-noise ratio, using matched filters,

$$\frac{a^2 [S_n, S_n]^2}{[R_c(S_n) S_n, S_n] + \phi [S_n, S_n]} \quad (10)$$

converges as  $n$  goes to infinity to the maximal signal-to-noise ratio attainable:

$$a^2 \left( \frac{E}{\phi} \right) \quad (11)$$

corresponding to the case of no clutter. Except in the trivially singular case, there is no signal which actually attains Expression (11).

**Proof.** For each  $S$  in  $\mathcal{H}$ , let

$$f(S) = [R_c(S) S, S]$$

which defines a homogeneous polynomial functional of order four on  $\mathcal{H}$ . This functional is actually Frechet differentiable. If we denote the gradient or Frechet derivative by  $L(\cdot)$ , so that

$$\frac{d}{d\lambda} f(S + \lambda H) \Big|_{\lambda=0} = [L(S), H], \quad H \in \mathcal{H}$$

then  $L(\cdot)$  is a homogeneous polynomial operator of order three, mapping  $\mathcal{H}$  into  $\mathcal{H}$ . Moreover, under Condition (9), it has the "compactness" or "complete continuity" property that it takes weakly convergent sequences in  $\mathcal{H}$  into

strongly convergent sequences. Also noting that

$$\begin{aligned}\left. \frac{d}{d\lambda} f(S + \lambda S) \right|_{\lambda=0} &= [L(S), S] \\ &= \left. \frac{d}{d\lambda} (1 + \lambda)^4 \right|_{\lambda=0} f(S) \\ &= 4f(S)\end{aligned}$$

we have that

$$[L(S), S] = 4f(S) \quad (12)$$

and, in particular, the right-hand side is non-negative since  $(R_c S, S) \geq 0$ . Now let  $S_n$  be any sequence of elements in  $\mathcal{H}$  such that

$$[S_n, S_n] = E$$

and such that  $S_n$  converges weakly to zero. (Of course, it is trivial that such sequences do exist in  $\mathcal{H}$ .) Then we know that  $L(S_n)$  converges strongly to zero, and, hence, that

$$\lim [L(S_n), S_n] = 0 = \lim f(S_n)$$

by Eq. (12). Hence, it follows that Expression (10) converges to Expression (11). If Expression (11) is actually attained by some signal, then for that signal we must, of course, have Eq. (8) holding; or, we must have trivial singularity. Also, for any sequence  $S_n$  for which Expression (10) converges to Expression (11), we must have that

$$\lim f(S_n) = 0$$

or, by Eq. (12),

$$\lim [L(S_n), S_n] = 0$$

But since

$$[S_n, S_n] = E$$

(i.e.,  $S_n$  is bounded), we can always extract a subsequence which converges weakly to an element  $S_0$ , say; and since  $L(S_n)$  converges strongly to  $L(S_0)$ , we must have

$$[L(S_0), S_0] = 0 = f(S_0)$$

If  $S_0$  is not zero, we have the trivially singular case.

**Remark 1.** It can be shown that for any (constant energy) sequence  $S_n$ , converging weakly to zero, the bandwidth occupancy must increase indefinitely with increasing  $n$ .

**Remark 2.** The sequence  $S_n$  in Theorem 1 can also be obtained by making use of a result due to Lusternik (Ref. 4), according to which we can always obtain a sequence of non-negative eigenvalues  $\lambda_n$  and corresponding eigenfunctions  $S_n$  for the operator  $L(\cdot)$  by virtue of its asserted properties:

$$L(S_n) = \lambda_n S_n, \quad [S_n, S_n] = E > 0$$

where  $\lambda_n$  converges to zero.

**Remark 3.** The square integrability Condition (9) is essential for the validity of Theorem 1. Here is a simple counter example to show this. Thus, let

$$R_c(S)W = [S, S]W$$

corresponding to suitable delta-functions in the clutter covariance. Then Expression (5) becomes

$$\frac{a^2 S, W^2}{[S, S][W, W] + \phi[W, W]} \leq \frac{a^2 [S, S]}{[S, S] + \phi}$$

and the right side is attained for  $S = W$ , and the maximal value

$$\frac{a^2 E}{E + \phi}$$

is independent of wave-shape, provided only that  $[S, S] = E$ .

### 3. Stationary Fast-Fading and Band-Limited Signals

We shall consider the special case where the fast-fading covariance is assumed stationary and, more specifically, we have

$$\begin{aligned}r_1(t, s) &= \mathbf{E}(\overline{N_c(t+L)} N_c(s+L)) \\ &= r_c(t-s) \int_0^T \overline{S(\sigma)} S(s-t+\sigma) d\sigma\end{aligned}$$

where  $r_c(\cdot)$  is a (complex) covariance function, while

$$r_2(t, s) = \mathbf{E}(N_c(t+L) N_c(s+L)) = 0$$

This was the model assumed in Ref. 1.

We have seen in Subsection 2 that the signal design problem is degenerate if no constraints are placed on the signals other than an upper bound on the energy. We

have usually additional constraints, however, and a natural constraint is that the allowed bandwidth occupancy be limited. Here we shall consider the case where the signals are strictly band-limited to a finite frequency interval  $I$ , so that the complex envelope is given by

$$S(t) = \int_I \psi(f) \exp 2\pi i f t df$$

because the complex analytic signal  $\psi(f)$  actually vanishes for negative  $f$ , but no specific use will need to be made of this fact. Now, of course, the signal cannot be time-limited; hence, we assume that the time-bandwidth product ( $2\pi IT$ ) is so large that we can effectively replace

$$2[R_c(S)S, S] = \int_0^T \int_0^T \overline{S(t)} r_c(t-s) r(t-s) S(s) ds dt$$

by

$$\int_0^\infty \int_0^\infty \overline{S(t)} r_c(t-s) r(t-s) S(s) ds dt \quad (13)$$

and generally  $T$  by infinity, where we have used the notation

$$r(t-s) = \int_0^T \overline{S(\sigma)} S(s-t+\sigma) d\sigma \quad (14)$$

and there is no need to specify "real part" since the double integral is already non-negative. The Functional (5) can now be expressed entirely in terms of frequency transforms, using

$$[S, W] = \operatorname{Re} \int_I \psi(f) \overline{\psi_w(f)} df$$

$$[R_c(S)W, W] = \int_I P(f) |\psi_w(f)|^2 df$$

where

$$\int_I \psi_w(f) \exp 2\pi i f t df = W(t)$$

$$P(f) = \int_I P_c(f-f') |\psi(f')|^2 df'$$

$$P_c(f) = \frac{1}{2} \int_{-\infty}^{\infty} r_c(t) \exp(-2\pi i f t) dt$$

The signal-to-noise ratio (Expression 5) then becomes<sup>7</sup>

$$\frac{\left( \operatorname{Re} \int_I \psi(f) \overline{\psi_w(f)} df \right)^2}{\int_I (P(f) + \phi) |\psi_w(f)|^2 df} \quad (15)$$

The maximum possible signal-to-noise ratio is the supremum of this functional over the class of functions  $S(\cdot), W(\cdot)$  such that

$$\int_I |\psi_w(f)|^2 df < \infty, \quad \int_I |\psi(f)|^2 df \leq E$$

As before, for a given  $S(\cdot)$ , there is an optimal  $W(\cdot)$  which maximizes Expression (15); in fact, this is given by

$$\psi_w(f) = \frac{\psi(f)}{P(f) + \phi}$$

and the corresponding value of Expression (15) is obviously given by

$$\frac{\int_I |\psi(f)|^2 df}{\int_I \phi + P(f)} \quad (16)$$

It is obvious that the maximum of Expressions (15) and (16) can be confined to the class of signals having the maximum energy  $E$ . As in Subsection 2, the supremum of Expression (16) is attained in the class of matched-filter receivers, subject only to the mild assumption that  $P_c(f)$  is continuous in  $f$  on finite intervals. Thus:

## THEOREM 2

*There exists a Lebesgue-Stieltjes measure on  $dP_0$  on  $I$  such that*

$$\int_I \hat{P}_c(f-f') dP_0(f') = \text{constant}, \quad f \in I \quad (17)$$

(i.e., with respect to  $dP_0$ )

$$\int_I dP_0(f) = E \quad (18)$$

where

$$2\hat{P}_c(f-f') = P_c(f-f') + P_c(f'-f)$$

<sup>7</sup>We have set  $a = 1$  since it plays no role in the design problem.

and

$$\sup \frac{\left( \int_I |\psi(f)|^2 df \right)^2}{\int_I (P(f) + \phi) |\psi(f)|^2 df} = \frac{E}{\phi + \gamma E} \quad (19)$$

$$= \sup \int_I \frac{|\psi(f)|^2 df}{\phi + P(f)}$$

where  $\gamma E$  is the constant in Eq. (17), and both suprema are taken over functions  $\psi(\cdot)$  such that

$$\int_I |\psi(f)|^2 df = E \quad (20)$$

**Proof.** Due to the length of the proof, it will not be included.

**Remark.** It may be noted that Eq. (17) has the important interpretation that the spectral density of the fast-fading component is a constant across the prescribed band for the optimal signal.

#### 4. A Communication System Application

Although the signal design problem has so far been considered in the radar or sonar context, we shall now indi-

cate how the results apply to the problem of coherent PCM biphase transmission of signals from a spacecraft in the vicinity of a planet. Here, the fast-fading component is due to surface scattering from the planet.<sup>8</sup> See Ref. 6 for a more complete description. Thus, let  $F(t)$  denote the transmitted complex envelope, so that the received signal in a given band can be expressed (cf Eq. 1):

$$X(t) = F(t) + N(t) + N_c(t), \quad 0 < t < T$$

where  $T$  is the signaling interval for each transmitted digit in the coherent biphase PCM system assumed, and  $S(t)$  is the modulation signal waveform:

$$F(t) = \sum_{-\infty}^{\infty} a_n S(t - nT), \quad S(t) = 0, t < 0, t > T$$

where the  $\{a_n\}$  is a random sequence of real numbers with

$$|a_n| = 1$$

$N(t)$  is the ambient noise (white gaussian), and  $N_c(t)$  is the clutter component

$$N_c(t) = \int_0^\infty K(t, s) F(t - s) ds = \int_0^t K(t, s) a_0 S(t - s) ds + \sum_1^\infty \int_0^T a_n K(t, t + nT + \sigma) S(\sigma) d\sigma$$

where  $K(t, s)$  has the same interpretation as in Subsection 2. We may note that the main difference from the radar case is that we have to account for the fast-fading contribution due to the signal in the previous bands. However, we note that we can still write  $N_c(t)$  in the form

$$N_c(t) = \int_0^t K(t, t - \sigma) a_0 S(\sigma) d\sigma + \sum_1^\infty a_n \int_0^T K(t, t + nT + \sigma) S(\sigma) d\sigma = \int_0^T \hat{K}(t, t - \sigma) S(\sigma) d\sigma, \quad 0 < t < T$$

where  $\hat{K}(t, s)$  can be read off from the previous line with the convention that it is zero for  $s$  negative. The similarity to Eqs. (1) and (2) is now clear. The signal design problem here is again the choice of the waveform  $S(t)$  in order to maximize detection probability. To see that the results of the previous subsections apply, we have only to note that, even though the clutter  $N_c(t)$  is not necessarily gaussian, as long as only linear operations are used,

the detection problem is equivalent to maximizing the signal-to-noise ratio based on the test statistic

$$\text{Re} \int_0^T X(t) \overline{W(t)} dt$$

<sup>8</sup>Lindsey, W. C., *Performance of Correlation Receivers in the Reception of Digital Signals Over Randomly Time-Varying Channels*, 1967 (JPL internal document).

Thus, we have the same mathematical problem as before, since we have seen that the optimal system can be sought in the class of "matched filter receiver" systems, the signal design problem being "degenerate" if there are no bandwidth constraints. The advantage of matched filter systems is quite apparent since the fast-fading return statistics depend also on the signal statistics over many bands and hence are difficult to obtain. When the stationarity assumption of *Subsection 3* is valid, we have the useful result that the optimal signal will make the fast-fading return have constant spectral density across the prescribed band.

### References

1. Balakrishnan, A. V., "Signal Design for a Class of Clutter Channels," *IEEE Trans. Inform. Theory*, pp. 170-173, Jan. 1968.
2. Balakrishnan, A. V., "Estimation and Detection Theory for Multiple Stochastic Processes," *J. Math. Anal. Appl.*, pp. 386-410, Dec. 1960.
3. Kailath, T., "Some Results on Singular Detection," *Inform. Contr.*, pp. 130-152, Apr. 1966.
4. Lusternik, L. A., "The Topology of the Calculus of Variations in the Large," Vol. 16, *Am. Math. Soc. Transl.*, 1966.

## G. Combinatorial Communications: Error-Correcting Prefixes for Establishing Word Synchronization, R. J. McEliece

### 1. Introduction

We shall present in this article a description of a simply implementable technique for establishing command word synchronization on the *Viking 1975* mission. This technique is quite general and is applicable in other situations, including future space missions.

The situation, briefly, is this: binary (0-1) commands are to be sent to the *Viking* spacecraft periodically, after quiet periods during which the idle sequence  $\cdots 0000 \cdots$  has been transmitted. The object is to establish synchronization of the command words as soon as the command is received, at least with very high probability. The space channel causes random bit errors with probability  $10^{-5}$ , and because the lock response time of the bit synchronizer is four bit periods, there may occur bursts of errors between lock decisions, although the probability of such bursts has not been accurately determined as yet.

The technique we wish to suggest for this situation is the use of a relatively short *prefix*, or word start signal,

which is to be transmitted prior to the transmission of commands. We shall discuss in the next subsection the selection of the appropriate prefix.

Actually, the situation on *Viking 1975* is at first glance somewhat more complicated. Because the bit synchronizer locks to transitions in the data stream, the best idle sequence is  $\cdots 10101010 \cdots$  and, in fact, it is planned to use this idle sequence, rather than all *zeros*. In addition, the bit synchronizer only recognizes the data up to complementation; e.g., there is no way the spacecraft can internally distinguish the sequence  $101010 \cdots$  from  $010101 \cdots$ . However, as R. C. Tausworthe has pointed out, if the commands are added modulo 2 to the idle sequence and then transmitted, and if the spacecraft adds a locally generated idle sequence  $\cdots 010101 \cdots$  to the received sequence, these complications disappear. For, whether the local idle sequence  $010101 \cdots$  is added to the received idle sequence in or out of phase, the resulting stream is a *constant* signal, which the spacecraft may *assume* is all *zeros*. When this is done, the transmitted data will be received directly. However, at the end of *Subsection 3* we shall also consider the selection of a prefix for *Viking 1975* if the data is transmitted directly, i.e., not mod 2 added to the idle stream. It will be seen that Tausworthe's observation saves 2 or 3 bits of prefix, at the cost of the extra hardware needed to do the mod-2 addition.

### 2. A General Formulation of the Problem

Thus, we shall consider the problem of establishing word synchronization on a noisy binary channel in which the idle sequence is all *zeros*. We shall solve this problem by transmitting a fixed "prefix" sequence which is recognized by the decoder and which signals the start of data transmission. To ensure the accurate acquisition of word synchronization, the prefix must be carefully chosen. First of all, it must be chosen so that it is very unlikely that the noisy channel can corrupt the all *zeros* idle sequence so badly that the prefix occurs in the received sequence before it was sent. More generally, if the prefix is  $P = p_1 p_2 \cdots p_n$ , each of the shifts  $00 \cdots 0 p_1 p_2 \cdots p_k$  must be sufficiently unlike  $P$  so that it is not likely that the decoder will recognize the prefix prematurely.

One drawback to this approach is that if the channel corrupts the prefix *itself* so that it is *not* recognized by the decoder, the decoder will continue its search for the prefix within the data itself; and unless either the prefix is quite long, or the data itself is structured to prevent the false recognition of the prefix in the data (as was done

on *Mariners VI and VII*), this event will be much more probable than the premature recognition of the prefix. A modification of this strategy, which overcomes the problem of not recognizing the prefix as it comes in, is to design the decoder so it will recognize the prefix even if it *has* been somewhat garbled by the channel. This strategy will greatly reduce the probability of not recognizing the prefix, and so also the chance of accidentally locking to part of the data that looks like the prefix. The next subsection contains a quantitative discussion of the choice of the prefix, if this strategy is to be used.

### 3. The Selection of the Prefix

Let  $P = p_1 p_2 \cdots p_n$  be a binary sequence (prefix) of length  $n$ . Define

$$d(P) = \min_{k \geq 1} d(P, P^k)$$

where  $d$  is hamming distance and

$$P^k = \underbrace{00 \cdots 0}_{\leftarrow k \rightarrow} p_1 p_2 \cdots p_{n-k}$$

Thus, it takes at least  $d(P)$  bit errors in the transmitted stream

$$\cdots 00 \cdots 0 p_1 p_2 \cdots p_n$$

to cause the prefix  $P$  to occur in the received stream prematurely. Let us call  $P$  a  $t$ -error correcting prefix if  $d(P) \geq 2t + 1$ . The following theorem justifies the terminology.

**THEOREM 1.** *If the decoder is designed to recognize a pattern  $Q$  of  $n$  bits as the prefix  $P$ , if and only if  $Q$  differs from  $P$  in  $\leq t$  bits, a necessary and sufficient condition that no pattern of  $t$  or fewer errors in any block of  $n$  consecutive transmitted bits causes the decoder to make the wrong decision on that block is that  $P$  be a  $t$ -error correcting prefix.*

**Proof.** If  $d(P) \geq 2t + 1$ , then  $d(P, P^k) \geq 2t + 1$  for each  $k \geq 1$ . Thus,  $t$  or fewer errors in any of the blocks  $P^k$  cannot distort  $P^k$  to within hamming distance  $t$  of  $P$  itself. Conversely, if  $d(P) \leq 2t$ , there will be some  $k$  for which  $d(P, P^k) \leq 2t$ ; and for this  $k$ , some pattern of  $t$  or fewer errors in  $P^k$  will cause  $P^k$  to differ from  $P$  in  $t$  or fewer positions.

Theorem 1 gives a criterion by which the prefix  $P$  should be chosen if the channel causes *independent* errors; but if the channel causes *bursts* of errors, a somewhat more general criterion is needed. To discuss the selection of prefixes which can correct *bursts* of errors, let us define the *burst- $b$  distance* between two binary  $n$ -tuples,

$$P = (p_1, p_2, \cdots, p_n)$$

and

$$Q = (q_1, q_2, \cdots, q_n)$$

$d^{(b)}(P, Q)$ , as the least integer  $d$  for which it is possible to select indices  $0 < i_1 < i_2 < \cdots < i_d \leq n$  such that  $P$  and  $Q$  agree except (possibly) in the positions

$$i_j + k, j = 1, 2, \cdots, d; k = 0, 1, \cdots, b - 1$$

Finally, define

$$d^{(b)}(P) = \min_{k \geq 1} d^{(b)}(P, P^k)$$

**THEOREM 2.** *If the decoder is designed to recognize a pattern  $Q$  of  $n$  bits as the prefix  $P$ , if and only if  $Q$  agrees with  $P$  except for at most  $t$   $b$ -bursts, a necessary and sufficient condition that no pattern of  $t$  or fewer  $b$ -bursts in any block of  $n$  consecutive transmitted bits causes the decoder to make the wrong decision on that block is  $d^{(b)}(P) \geq 2t + 1$ .*

**Proof.** The proof of Theorem 1 carries over, *mutatis mutandis*.

Table 1 gives a list of the shortest prefixes for which  $d^{(b)}(P)$  has a given value; they were found by computer search. The case  $b = 1$  is the case of Theorem 1.

Finally, Table 2 gives a list of the shortest prefixes which are appropriate for use with the *Viking 1975* commands if the prefix and commands are sent directly; i.e., not mod-2 added to the 101010  $\cdots$  idle sequence, and if only *one* burst is to be corrected. (It will certainly be sufficient to correct a single burst on the *Viking 1975* mission, since the probability of *two* or more such bursts in a short span is conservatively estimated at less than  $10^{-8}$ .) (The duration of these rare bursts is questionable.) It should be observed that these prefixes are at most three bits longer than the corresponding prefixes in Table 1.

**Table 1. Some of the shortest binary prefixes which have a fixed value of  $d^{(b)}$  (octal notation<sup>a</sup>)**

$d$	$b = 1$		$b = 2$		$b = 3$		$b = 4$		$b = 5$		$b = 6$		$b = 7$		$b = 8$	
	$n$	Prefix	$n$	Prefix	$n$	Prefix	$n$	Prefix	$n$	Prefix	$n$	Prefix	$n$	Prefix	$n$	Prefix
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	3	6	4	13	5	23	6	43	7	103	8	203	9	403	10	1003
3	5	32	7	123	9	456	11	2126	13	10246	15	40506	17	201206	19	1002406
4	7	134	10	1236	13	10766	15	42547	18	412307						
5	9	571	12	5342	16	117534	20	2115274								
6	11	2670	15	53543	20	2232147										
7	12	7531	18	534664												
8	14	35330														
9	16	165542														
10	18	734264														
11	20	3553426														

<sup>a</sup>The  $d = 3$  row corresponds to the shortest single burst- $b$  error-correcting prefixes.

**Table 2. Some of the shortest single burst- $b$  error-correcting prefixes for use on Viking 1973 if data is not mod-2 added to 10101010 idle sequence<sup>a</sup>**

$b = 1$		$b = 2$		$b = 3$		$b = 4$		$b = 5$		$b = 6$		$b = 7$	
8	362	9	751	11	3062	13	14742	15	74315	17	367251	19	1701475

<sup>a</sup> $d = 3$ .

## H. Combinatorial Communications: Some Results on the Capacity of Graphs, R. J. McEliece, R. P. Stanley, and H. Taylor

### 1. Introduction

Let  $C$  be a finite discrete memoryless channel which is specified by its transition probability matrix  $[p(j|i)]$ , where  $p(j|i)$  is the probability that input letter  $i$  will be received as output letter  $j$  ( $i = 1, 2, \dots, I$ ;  $j = 1, 2, \dots, J$ ). In 1956, Shannon (Ref. 1) defined the *zero-error capacity* of such a channel as the least upper bound of rates at which it is possible to transmit information at zero-error probability. Shannon observed that the *adjacency graph* of  $C$  is fundamental to the study of zero-error capacity. The adjacency graph  $G$  has  $I$  vertices  $v_1, v_2, \dots, v_I$ ;  $v_i$  is connected to  $v_{i'}$  if there is an output letter  $j$  such that  $p(j|i)$  and  $p(j|i')$  are both nonzero. That is,  $v_i$  and  $v_{i'}$  are connected in  $G$  if the input letters  $i$  and  $i'$  can be confused by the channel. The importance of the adjacency graph may be seen as follows:

Suppose  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  and  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  are two codewords in a zero-error probability block code of length  $n$  from  $C$ . Now  $\mathbf{x}$  and  $\mathbf{y}$  can be confused by the channel only if  $x_i$  and  $y_i$  can be confused for each  $i$ ; but this is equivalent to saying that  $\mathbf{x}$  and  $\mathbf{y}$  are connected in the graph  $G^n$ . [If  $G$  is a graph, the direct power  $G^n$  has as vertices the set of  $n$ -tuples  $(v_1, v_2, \dots, v_n)$ , where  $v_i$  are vertices of  $G$ ;  $(v_1, v_2, \dots, v_n)$  and  $(v'_1, v'_2, \dots, v'_n)$  are connected if and only if for each  $i$  either  $v_i = v'_i$  or  $v_i$  and  $v'_i$  are connected in  $G$ .] Thus, the number of words in the largest error-free code of length  $n$  from  $C$  is the largest number of vertices in  $G^n$ , no two of which are adjacent. Berge (Ref. 2) defines the *coefficient of internal stability* of any graph  $G$ ,  $\alpha(G)$ , as the largest number of vertices of  $G$  which may be chosen such that no two are adjacent. Hence, using Berge's notation, the zero-error capacity of  $C$  is

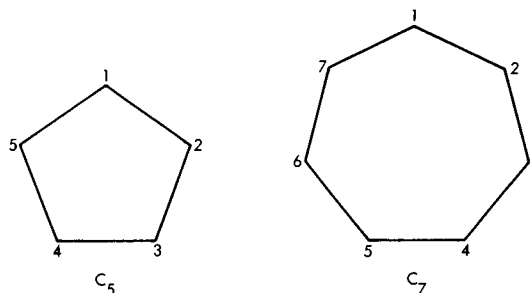
$$\sup_n \frac{1}{n} \log \alpha(G^n)$$

(And, in fact, it is not hard to see that we may replace "sup" by "lim".) This leads us to define the *capacity* of any finite undirected graph  $G$  as

$$\text{cap}(G) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha(G^n)$$

It turns out that for most graphs  $G$ ,  $\text{cap}(G) = \log \alpha(G)$ ; but for the few which have  $\text{cap}(G) > \log \alpha(G)$ ,  $\text{cap}(G)$  is unknown! It is the object of this article to study the

functions  $\alpha$  and  $\text{cap}$ , especially as applied to the so-called *odd cycle graphs*  $C_{2m+1}$ , which have an odd number of vertices  $v_1, v_2, \dots, v_{2m+1}$  and for which  $v_i$  is connected to  $v_j$  if and only if  $i - j \equiv \pm 1 \pmod{2m+1}$ .  $C_5$  and  $C_7$  are illustrated below:



It is very easy to show that

$$\log \frac{p-1}{2} \leq \text{cap}(C_p) \leq \log \frac{p}{2}$$

We shall be able to increase this lower bound for all odd  $p \geq 5$ ; in particular, we shall show

$$\alpha(C_p^2) = \left\lfloor \frac{p^2 - p}{4} \right\rfloor$$

for all odd  $p$ , and that

$$\text{cap}(C_p) > \frac{1}{2} \log \alpha(C_p^2)$$

for infinitely many  $p$ , including  $p = 7$  and  $p = 9$ .

### 2. A Useful Result About $\alpha$

We begin with some definitions; throughout  $G$  is a finite undirected graph. A *clique* in  $G$  is a set of vertices of  $G$  such that every pair is connected in  $G$ . A *brouhaha* is a set of vertices, no two of which are connected. The *dual graph*  $\bar{G}$  of  $G$  has the same vertices as  $G$ , but  $v$  and  $v'$  are connected in  $\bar{G}$  if and only if they are not connected in  $G$ . Thus, a clique in  $G$  is a brouhaha in  $\bar{G}$ , and conversely. Finally, if  $H$  is another graph, and if each vertex of  $G$  is part of exactly  $r$  subgraphs of  $G$  isomorphic to  $H$  for some fixed  $r$ , we say that  $G$  is *H-regular*. (Thus, ordinary regularity is a special case of  $H$ -regularity, where  $H$  is the graph consisting of two connected points.)

**THEOREM 1.** If  $G$  is  $H$ -regular, then  $\alpha(G)/|G| \leq \alpha(H)/|H| \cdot (|G|)$  is the number of vertices in  $G$ .

**Proof.** Suppose  $X$  is the largest brouhaha in  $G$ ;  $|X| = \alpha(G)$ . Then for each subgraph  $H'$  of  $G$  which is isomorphic to  $H$ , the vertices of  $H'$  which are members of  $X$  form a brouhaha in  $H'$ , and so  $|X \cap H'| \leq \alpha(H') = \alpha(H)$ . Since each vertex of  $G$  is part of  $r$  copies of  $H$  and each copy of  $H$  involves  $|H|$  vertices of  $G$ , there are  $|G| \cdot r / |H|$  copies of  $H$  altogether; thus,

$$\alpha(G) \cdot r = \sum_{\text{copies of } H} |X \cap H'| \leq \alpha(H) \cdot \left( \frac{|G| \cdot r}{|H|} \right)$$

and the theorem follows.

**Corollary 1 (the sphere-packing bound).** If  $G$  is  $H$ -regular and  $H$  is a clique, then  $\alpha(G) \leq |G|/|H|$  and

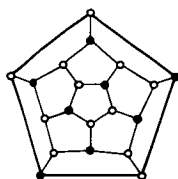
$$\text{cap}(G) \leq \log \left( \frac{|G|}{|H|} \right)$$

**Proof.** The first part follows, since  $\alpha(H) = 1$  if  $H$  is a clique. The second part follows from the fact that  $G^n$  is  $H^n$ -regular for all  $n$  and  $H^n$  is a clique.

**Corollary 2.** More generally, if  $G$  is  $H$ -regular,

$$\text{cap}(G) \leq \text{cap}(H) + \log \left( \frac{|G|}{|H|} \right)$$

Let us give one example of the use of Theorem 1. Suppose  $G$  is the graph of the regular dodecahedron:



The eight distinguished vertices show that  $\alpha(G) \geq 8$ ; on the other hand,  $G$  is  $C_5$ -regular ( $C_5$  is the pentagon). Thus, by Theorem 1  $\alpha(G) \leq (20/5) \cdot \alpha(C_5) = 8$ , and so  $\alpha(G) = 8$ .

### 3. The Graphs $C_p$

In this subsection, we shall restrict our attention to the graphs  $C_p$  for odd  $p \geq 5$  mentioned in Subsection 1. If the vertices are numbered  $1, 2, \dots, p$ , the vertices  $1, 3, 5, \dots, p-2$  are a brouhaha, so that

$$\alpha(C_p) \geq \frac{p-1}{2}$$

On the other hand,  $C_p$  is  $C_2$ -regular, so that by Theorem 1  $\alpha(C_p) \leq p/2$ . Thus,  $\alpha(C_p) = (p-1)/2$ , and so

$$\text{cap}(C_p) \geq \log \frac{p-1}{2}$$

Also, by Corollary 2 to Theorem 1,  $\text{cap}(C_p) \leq \log p/2$ . We shall not be able to decrease this upper bound for any  $p$ , but Theorem 2 increases the lower bound for all odd  $p \geq 5$ .

**THEOREM 2.**  $\alpha(C_p^2) = \lfloor (p^2 - p)/4 \rfloor$ , and so

$$\text{cap}(C_p) \geq \frac{1}{2} \log \left\lfloor \frac{1}{4} (p^2 - p) \right\rfloor$$

**Proof.** We first show that  $\alpha(C_p^2) \leq \lfloor (p^2 - p)/4 \rfloor$ . This follows directly from Theorem 1, since  $C_p^2$  is  $C_p \times C_2$ -regular, and

$$\alpha(C_p \times C_2) = \alpha(C_p) = \frac{p-1}{2}$$

To show that  $\alpha(C_p^2) \geq \lfloor (p^2 - p)/4 \rfloor$  as well, we must explicitly exhibit a brouhaha in  $C_p^2$  of that size:

For  $p \equiv 1 \pmod{4}$ , say  $p = 4a + 1$ :

$$(t, 2t + 4s) \begin{cases} t = 0, 1, \dots, p-1 \\ s = 0, 1, \dots, a-1 \end{cases}$$

For  $p \equiv 3 \pmod{4}$ , say  $p = 4a + 3$ :

$$(2s, 2t + s) \begin{cases} s = 0, 1, \dots, 2a+1 \\ t = 0, 1, \dots, a-1 \end{cases}$$

and

$$(2s+1, 2t+s+2a+1) \begin{cases} s = 0, 1, \dots, 2a \\ t = 0, 1, \dots, a \end{cases}$$

We omit the straightforward but tedious verification that these sets of vertices do form brouhahas and regard Theorem 2 as proved.

Theorem 2 shows that for all odd  $p \geq 5$ ,

$$\text{cap}(C_p) > \log \alpha(C_p)$$

Indeed, combining our results, we have shown that for  $p \geq 5$ ,

$$\frac{1}{2} \log \left\lfloor \frac{1}{4} (p^2 - p) \right\rfloor \leq \text{cap}(C_p) \leq \log \frac{p}{2}$$

and for  $p = 5$ , that is where matters have stood since Shannon's original paper. In *Subsections 4 and 5*, we will show that this lower bound can be improved for infinitely many  $p$ , including  $p = 7$  and  $p = 9$ .

#### 4. Good Packings in $C_p^n$ for $n \geq 3$

**THEOREM 3.**  $\alpha(C_p^n) \leq \frac{p}{2} \alpha(C_p^{n-1})$ .

*Proof.* Immediate from Theorem 1, since  $C_p^n$  is  $C_p^{n-1} \times C_2$ -regular, and  $\alpha(C_p^{n-1} \times C_2) = \alpha(C_p^{n-1})$ .

*Corollary.*  $\alpha(C_p^n) \leq \frac{(p^n - p^{n-1})}{2^n}$ .

*Proof.* From Theorem 3 (or Theorem 2),

$$\alpha(C_p^2) \leq \frac{p^2 - p}{4}$$

The corollary follows from Theorem 3 by induction on  $n$ .

Although for fixed  $p$ , as  $n$  increases the upper bound of the corollary is doubtless very crude; for fixed  $n$  and large  $p$ , it is probably very good. In particular, we present the following conjecture:

For all  $p \geq 2^n + 1$ ,  $\alpha(C_p^n) = \frac{p}{2} \alpha(C_p^{n-1})$

And while we will not be able to prove this conjecture (except for  $n = 2$ ), we will be able to prove Theorem 4, which is related:

**THEOREM 4.** If  $p = 2^n + 1$ ,

$$\alpha(C_p^n) = \frac{p^n - p^{n-1}}{2^n} = p^{n-1}$$

*Proof.* We have seen that

$$\alpha(C_p^n) \leq \frac{p^n - p^{n-1}}{2^n}$$

Thus, it remains to exhibit a brouhaha of size  $p^{n-1}$ . We claim the following set will do:

$$(x_1, x_2, \dots, x_{n-1}, x_n)$$

where  $x_1, \dots, x_{n-1}$  are arbitrary and

$$x_n = 2x_1 + 4x_2 + \dots + 2^{n-1}x_{n-1} \pmod{p}$$

To see this, suppose  $x = (x_1, \dots, x_n)$  and  $y = (y_1, \dots, y_n)$  are two vectors in the set, and that they are connected in  $C_p^n$ . Then

$$x - y = (x_1 - y_1, \dots, x_n - y_n) = (z_1, \dots, z_n)$$

has all coordinates congruent to 0, or  $\pm 1 \pmod{p}$ . But since

$$z_n = 2z_1 + \dots + 2^{n-1}z_{n-1}$$

we must show that if each  $z_i, i < n$  is 0,  $\pm 1 \pmod{p}$ , that  $2z_1 + \dots + 2^{n-1}z_{n-1}$  cannot be. Now let  $P$  be the set of indices  $i$  for which  $z_i = +1$ , and let  $M$  be the set for which  $z_i = -1$ . Then, if

$$2z_1 + \dots + 2^{n-1}z_{n-1} \equiv 0, \pm 1 \pmod{p}$$

we have a congruence of the form

$$\sum_{i \in P} 2^i \equiv \sum_{j \in M} 2^j \pmod{p}$$

(where either  $P$  or  $M$  has been intended to include 0, if necessary). But unless both sums are empty, they represent different integers of the range  $[0, 2^n - 1] = [0, p - 2]$  and so cannot be congruent  $\pmod{p}$ .

#### 5. Miscellaneous Results

We present three miscellaneous results concerning  $\alpha(C_p^n)$  and  $\text{cap}(C_p)$ .

**THEOREM 5.**  $\text{cap}(C_7) \geq \frac{3}{5} \log 7 > \frac{1}{2} \log 10$ .

*Proof.* It is easy to verify that the following set of  $7^3$  vertices is a brouhaha in  $C_7^3$ :

$$(x_1, x_2, x_3, 2x_1 + 2x_2 + 2x_3, 2x_1 + 4x_2 + 6x_3)$$

$x_1, x_2, x_3$  are arbitrary.

**THEOREM 6.**  $31 \leq \alpha(C_7^3) \leq 35$ .

**Proof.** The upper bound comes from Theorem 3 with  $p = 7$ ,  $n = 3$ . Here is a brouhaha of size 31:

5	3	1				7
		5	3	1		
3	1			5	3	1,5
	5	3	1			
1			5	3	1,5	3
5	3	1				
		5	3	1,6	4	2

It has often been conjectured that  $\alpha(C_5^n) = 5\alpha(C_5^{n-2})$  for  $n \geq 3$ ; this conjecture has been verified by exhaustive enumeration for  $n \leq 4$ . [Notice that it is sufficient to verify the conjecture for odd  $n$  by Theorem 3, since  $\alpha(C_5^n) \geq 5\alpha(C_5^{n-2})$ .]

Now, by a *systematic* brouhaha, we mean a brouhaha like the ones exhibited in Theorems 4 and 5, i.e., one of the form

$$(x_1, x_2, \dots, x_k, y_1, y_2, \dots, y_{n-k})$$

where the  $x_i$ 's vary freely over the integers (mod  $p$ ), and the  $y_i$ 's are uniquely determined by the  $x$ 's. Let us denote the size of the largest *systematic* brouhaha in  $C_p^n$  by  $\alpha_{sys}(C_p^n)$ . Note that  $\alpha_{sys}(C_5^2) = \alpha(C_5^2)$ .

**THEOREM 7.**  $\alpha_{sys}(C_5^n) = 5^{\lfloor n/2 \rfloor}$  for  $n \leq 12$ . (Thus, there is no systematic brouhaha which improves the bound  $\text{cap}(C_5) \geq \frac{1}{2} \log 5$  for  $n \leq 12$ .)

**Proof** If  $k = \lfloor n/2 \rfloor$ , then for any  $n$ , the set

$$(x_1, x_2, \dots, x_k, 2x_1, 2x_2, \dots, 2x_k, 0)$$

is a brouhaha of size  $5^k$  in  $C_5^n$ . (The 0 coordinate is only present for odd  $k$ .) From  $\alpha(C_5^3) = 10$ ,  $\alpha(C_5^4) = 25$ , and using

$$5\alpha(C_5^{n-2}) \leq \alpha(C_5^n) \leq \frac{5}{2}\alpha(C_5^{n-1})$$

we obtain the following sequence of upper and lower bounds on  $\alpha(C_5^n)$ :

$n$	Lower bound	Upper bound
5	50	62
6	125	155
7	250	387
8	625	967
9	1250	2417
10	3125	6042
11	6250	15105
12	15625	37762
13	31250	94405

Now, since the number of vertices in a systematic brouhaha is always a power of 5, and  $5^{\lfloor n/2 \rfloor}$  is the largest power of 5 less than the upper bound  $n \leq 12$ , the theorem follows. Finally, notice that  $5^7 = 78125 < 94405$ , so that it is conceivable that there is a systematic brouhaha for  $n = 13$  which would show  $\text{cap}(C_5) \geq 7/13 \log 5$ . However, if it turns out that  $\alpha(C_5^8) \leq 51$ , the above procedure would rule this out.

## References

1. Shannon, E. C., "The Zero-Error Capacity of a Noisy Channel," *IRE Trans. Inform. Th.*, IT-2, pp. 8-19, 1956.
2. Berge, C., *The Theory of Graphs*. Methuen & Co., Ltd., London, England, 1962.

## I. Combinatorial Communications: Negative Radix Conversion, S. Zohar

### 1. Introduction

In the common positional representation of numbers, negative numbers always present a special case. Thus, a machine that can add  $3 + 5$  has to go through a special sequence when the problem is to add  $3 + (-5)$ .

It has recently been pointed out<sup>9</sup> that a computer mechanization which is completely indifferent to the sign of a number can be built if, instead of the standard positive radix usually adopted in number representation, a negative radix is chosen.

<sup>9</sup>News item in *Electronics*, Vol. 40, No. 26, pp. 40-41, Dec. 25, 1967. The idea is credited to Mauritz P. de Regt.

Our purpose in this article is to develop algorithms for the transformation of the representation of a number as the radix is changed from  $|q|$  to  $-|q|$  and vice versa. We give the logic design of a device that will serially convert standard binary numbers to base  $-2$  representation. The range of numbers representable by a given number of digits is considered.

## 2. Conversion From Positive to Negative Base

Let  $q > 0$  be the base for positional number representation. Then an arbitrary number  $b$  is expressible as follows:

$$b = \sum_r \beta_r q^r \quad (S = +1, -1; \beta_r = 0, 1, \dots, q-1) \quad (1)$$

where  $r$  can range over both positive and negative integers.

Our problem is to find the following alternate representation of  $b$

$$b = \sum_r \beta'_r (-q)^r \quad (\beta'_r = 0, 1, \dots, q-1) \quad (2)$$

We note at the outset that since any base provides a unique representation, the set  $\beta'_r$  is unique. Consider now segregating Eq. (2) into even and odd powers:

$$b = \sum_{r \text{ even}} \beta'_r q^r - \sum_{r \text{ odd}} \beta'_r q^r \quad (3)$$

This suggests the usefulness of the following terminology: A number  $g$  is said to be *p-even* if it is expressible as

$$g = \sum_{r \text{ even}} \gamma_r p^r \quad (p \text{ integer; } \gamma_r = 0, 1, 2, \dots, |p|-1)$$

Similarly, a number  $g$  is said to be *p-odd* if it is expressible as

$$g = \sum_{r \text{ odd}} \gamma_r p^r \quad (\gamma_r = 0, 1, 2, \dots, |p|-1)$$

Returning now to Eq. (3), we denote

$$\left. \begin{aligned} \sum_{r \text{ even}} \beta'_r q^r &= e \\ \sum_{r \text{ odd}} \beta'_r q^r &= d \end{aligned} \right\} \quad (4)$$

Hence,

$$e - d = b \quad (5)$$

$$e + d = \sum_r \beta'_r q^r \quad (6)$$

where  $e$  is  $q$ -even,  $d$  is  $q$ -odd, and the range of  $r$  is the same as in Eq. (2). Obviously,  $e, d$  are unique since they result from a unique partitioning of the unique set  $\beta'_r$ .

Equations (5), (6), and (2) are the basis for the conversion algorithm. Given  $b$ , we decompose it according to Eq. (5). This is a straightforward process best illustrated by an example. Having obtained  $e, d$ , we add them. So far, all the representations have been in base  $q$ . Equation (6) tells us now that the result of the addition is a number whose representation in base  $q$  is

$$(\beta'_{q-1}, \dots, \beta'_1, \beta'_0)$$

But, according to Eq. (2), this is just the  $(-q)$  representation of  $b$ , which is what we are after.

Note that the addition  $(e + d)$  involves no computation since every non-zero digit of one term is matched to a zero digit in the other. Thus, we really have here just a merging of digits, and the main conversion effort is in the decomposition of  $b$ . In performing this decomposition, it is slightly more convenient to formulate Eq. (5) as an addition. Hence, we rephrase it

$$|b| + d = e \quad (S = 1) \quad (7)$$

$$|b| + e = d \quad (S = -1) \quad (8)$$

To illustrate the method, consider the conversion of the number 3749 to base  $-10$  (all numbers without a base index are base 10 representations).

We start by writing down all that we do know of the three numbers in the following scheme representing Eq. (7).

$$\begin{array}{rcccccccl} b = & 0 & 3 & 7 & 4 & 9 & & 03749 \\ & + & & & & & & \\ d = & 0 & \beta'_3 & 0 & \beta'_1 & 0 & & 07060 \\ e = & \beta'_4 & 0 & \beta'_2 & 0 & \beta'_0 & & 10809 \\ & & & & & & & \\ & & & & & & & e + d = 17869 \\ & & & & & & & b = 17869_{-10} \end{array}$$

The  $\beta'_i$  stand for the unknown digits. The zero digits in  $d, e$  satisfy the  $q$ -odd,  $q$ -even conditions.

Starting from the least significant digit, we find the unknown digits one by one:

$$\begin{aligned}\beta'_0 &= 9 + 0 = 9 \\ 4 + \beta'_1 &= 0 + 10(\text{carry}); \therefore \beta'_1 = 6 \\ \beta'_2 &= 1(\text{carry}) + 7 + 0 = 8 \\ 3 + \beta'_3 &= 0 + 10(\text{carry}); \therefore \beta'_3 = 7 \\ \beta'_4 &= 1(\text{carry}) + 0 + 0 = 1\end{aligned}$$

This is equivalent to  $b = 17869_{-10}$ .

For a negative number, the starting point is Eq. (8). The following is the algorithm for converting  $-3749$ :

$$\begin{array}{rcccccccl} -b = & 3 & 7 & 4 & 9 & & 3749 \\ + & & & & & & + \\ e = & 0 & \beta'_2 & 0 & \beta'_0 & & 0301 \\ d = & \beta'_3 & 0 & \beta'_1 & 0 & & 4050 \\ & & & & & e + d = & 4351 \\ & & & & & b = & 4351_{-10} \end{array}$$

### 3. Conversion From Negative to Positive Base

Compared to the previous case, this conversion is trivial. Given  $b$  in base  $-q$ , means that all  $\beta'_i$  are known. Hence, according to Eq. (6),  $e + d$  in base  $q$  is known. We extract  $e, d$  from  $e + d$  by a simple reversal of the merging process of Subsection 2. Finally,  $b$  (in base  $q$ ) is given by Eq. (5).

We illustrate this with the numbers of the previous examples. For  $b = 17869_{-10}$ , we get

$$\begin{aligned}b &= 17869_{-10} \\ e + d &= 17869 \\ e &= 10809 \\ &- \\ d &= 7060 \\ b &= 3749\end{aligned}$$

Similarly, for  $b = 4351_{-10}$ , we get

$$\begin{aligned}b &= 4351_{-10} \\ e + d &= 4351 \\ e &= 301 \\ &- \\ d &= 4050 \\ b &= -3749\end{aligned}$$

Before leaving this subject, a few words are due in explanation of the seeming asymmetry of the algorithm as regards the direction of conversion. It should be noted that though it was assumed in Subsection 2 that  $q > 0$ , this was done mainly for convenience in presentation and the basic equations are true regardless of the sign of  $q$ . Why then is the algorithm of this subsection different?

Ignoring for a moment the algorithm of Subsection 2, we maintain that the present algorithm is applicable to the conversion from base 10 to base  $-10$ . We have only to realize that, in the scheme of this subsection, we start by writing down  $b$  in the given base and, following that, all other representations are in the negative of this base. Applying this to the number 3749, we get

$$\begin{aligned}b &= 3749_{10} \\ e + d &= 3749_{-10} \\ e &= 709_{-10} \\ &- \\ d &= 3040_{-10} \\ b &= 17869_{-10}\end{aligned}$$

Note that we have here the subtraction of two numbers represented in a negative base. The algorithm for such a subtraction differs from the positive base algorithm in the treatment of "borrowing" and carry:

- (1) Compensation for "borrowing" is the *increase* of the "lending" digit by 1. Thus, suppose we have a digit  $m$  representing  $mq^n$  and would like to increase it to  $m + |q|$  by "borrowing." For negative  $q$ , we have the following identity proving our contention:

$$mq^n = (m + |q|)q^n + q^{n+1} \quad (9)$$

Incrementing a digit as shown here could cause a carry, which is handled next.

- (2) A carry adds  $|q| - 1$  to the next digit and 1 to the following digit. This is verified by the following identity, valid for negative  $q$ :

$$|q|q^n = (|q| - 1)q^{n+1} + q^{n+2} \quad (10)$$

The conversion of  $-3749$  is shown below:

$$\begin{aligned} b &= -3749_{10} \\ e + d &= -3749_{-10} \\ e &= -709_{-10} \\ &\quad - \\ d &= -3040_{-10} \\ b &= 4351_{-10} \end{aligned}$$

We see that there is, in fact, complete symmetry. The different form of the algorithm displayed in *Subsection 2* is just the result of trying to avoid the subtraction of negative base numbers.

It should be observed that a device that utilizes numbers in negative base representation might probably also have the capability of arithmetic operations with such numbers. This would favor the adoption of the present algorithm for the conversion of input data. However, the situation assumed above is not necessarily always the case since there are applications which do not involve a negative base arithmetic unit.

#### 4. A Serial Converter for a Binary Machine

We consider here the implementation of the algorithm of *Subsection 2* for a binary machine. We treat first the

case of positive  $b$  to which Eq. (7) applies:

$$|b| + d = e$$

Let us denote by  $\alpha_r$  the carry created in position  $r - 1$  and carried into position  $r$ . For even  $r$ , the bit configuration in  $b, d, e$  is

$$\begin{array}{r} b \\ + \\ d \\ \hline e \end{array} \quad \begin{array}{r} -\alpha_r \\ \beta_r \\ + \\ 0 \\ \hline \beta'_r \end{array}$$

Hence,  $\beta'_r$  will be 1 if  $\beta_r = 1$  and  $\alpha_r = 0$ , or if  $\beta_r = 0$  and  $\alpha_r = 1$ . That is,

$$\beta'_r = \beta_r \bar{\alpha}_r + \bar{\beta}_r \alpha_r \quad (11)$$

A carry into position  $r + 1$  will be created only if  $\beta_r = 1$  and  $\alpha_r = 1$ . Hence,

$$\alpha_{r+1} = \beta_r \alpha_r \quad (12)$$

When  $r$  is odd, the bit configuration is

$$\begin{array}{r} b \\ + \\ d \\ \hline e \end{array} \quad \begin{array}{r} -\alpha_r \\ \beta_r \\ + \\ \beta'_r \\ \hline 0 \end{array}$$

The logic equation for  $\beta'_r$  is still the same (Eq. 11). The generation of a carry, however, is different. With the  $e$ -bit being zero, no carry will be created if  $\beta_r = 0$  and  $\alpha_r = 0$ . Hence,

$$\left. \begin{aligned} \bar{\alpha}_{r+1} &= \bar{\beta}_r \bar{\alpha}_r \\ \therefore \alpha_{r+1} &= \beta_r + \alpha_r \end{aligned} \right\} \quad (13)$$

When  $b$  is negative,  $e$  and  $d$  change roles. This leads to the interchange of the carry equations. We summarize the logic equations:

$$\beta'_r = \beta_r \bar{\alpha}_r + \bar{\beta}_r \alpha_r = \overline{\beta_r \bar{\alpha}_r \bar{\beta}_r \alpha_r} \quad (14)$$

$$\left. \begin{aligned} \alpha_{r+1} &= \beta_r \alpha_r & \text{if } b > 0 \text{ and } r \text{ is even, or } b < 0 \text{ and } r \text{ is odd} \\ \alpha_{r+1} &= \beta_r + \alpha_r & \text{if } b > 0 \text{ and } r \text{ is odd, or } b < 0 \text{ and } r \text{ is even} \end{aligned} \right\} \quad (15)$$

To handle the carry, we introduce now the sign bit  $\sigma$ :

$$\left. \begin{array}{ll} \sigma = 0 & \text{for } S = 1 \\ \sigma = 1 & \text{for } S = -1 \end{array} \right\} \quad (16)$$

We assign a  $JK$  flip-flop ( $H$ ) to switch the logic between the two carry equations. We set the initial state of  $H$  (for  $r = 0$ ):

$$H_0 = \sigma \quad (17)$$

At each stepping of  $r$ ,  $H$  is toggled ( $H_1 = \bar{\sigma}$ ;  $H_2 = \sigma$ ,  $\dots$ ). Hence, both carry equations can now be combined into the following single recursion:

$$\alpha_{r+1} = \bar{H}_r \beta_r \alpha_r + H_r (\beta_r + \alpha_r) \quad (\alpha_0 = 0) \quad (18)$$

Modifying with the aim of *nand* implementation, we get

$$\alpha_{r+1} = \overline{\alpha_r H_r \alpha_r \beta_r H_r \beta_r} \quad (\alpha_0 = 0) \quad (19)$$

The overall logic diagram is shown in Fig. 6, which is a direct implementation of Eqs. (14) and (19). The main clock sequence controlling the device is  $CL$ , which may be referred to as the bit sequence. Each clock pulse generates a new bit of the converted number. The remaining two clock sequences are derived from  $CL$ .

Consider  $CL_1$ , the word sequence, first. For  $r \neq 0$ ,  $CL_1$  blocks transfer of information from the sign bit ( $\sigma$ ) flip-flop and establishes  $J = K = 1$ . Hence, flip-flop  $H$  will toggle whenever its clock input ( $CL$ ) is 1. For  $r = 0$ , the  $J, K$  gates implement the transfer of the  $\sigma$  bit to  $H$ , thus establishing  $H_0 = \sigma$ .

$CL_1$  also implements the initialization of the carry shift register. The direct connection of  $CL_1$  to the reset (clear) input of the  $\alpha_r$  flip-flop establishes  $\alpha_0 = 0$ . At the same time, the application of  $CL$ ,  $CL_1$  to the shift input, as shown, inhibits shifting when  $r = 0$ . Sequence  $CL_2$ , which drives the input shift register, is a replica of sequence  $CL$  with the omission of the two last bit pulses in each word. This is necessary since the converted

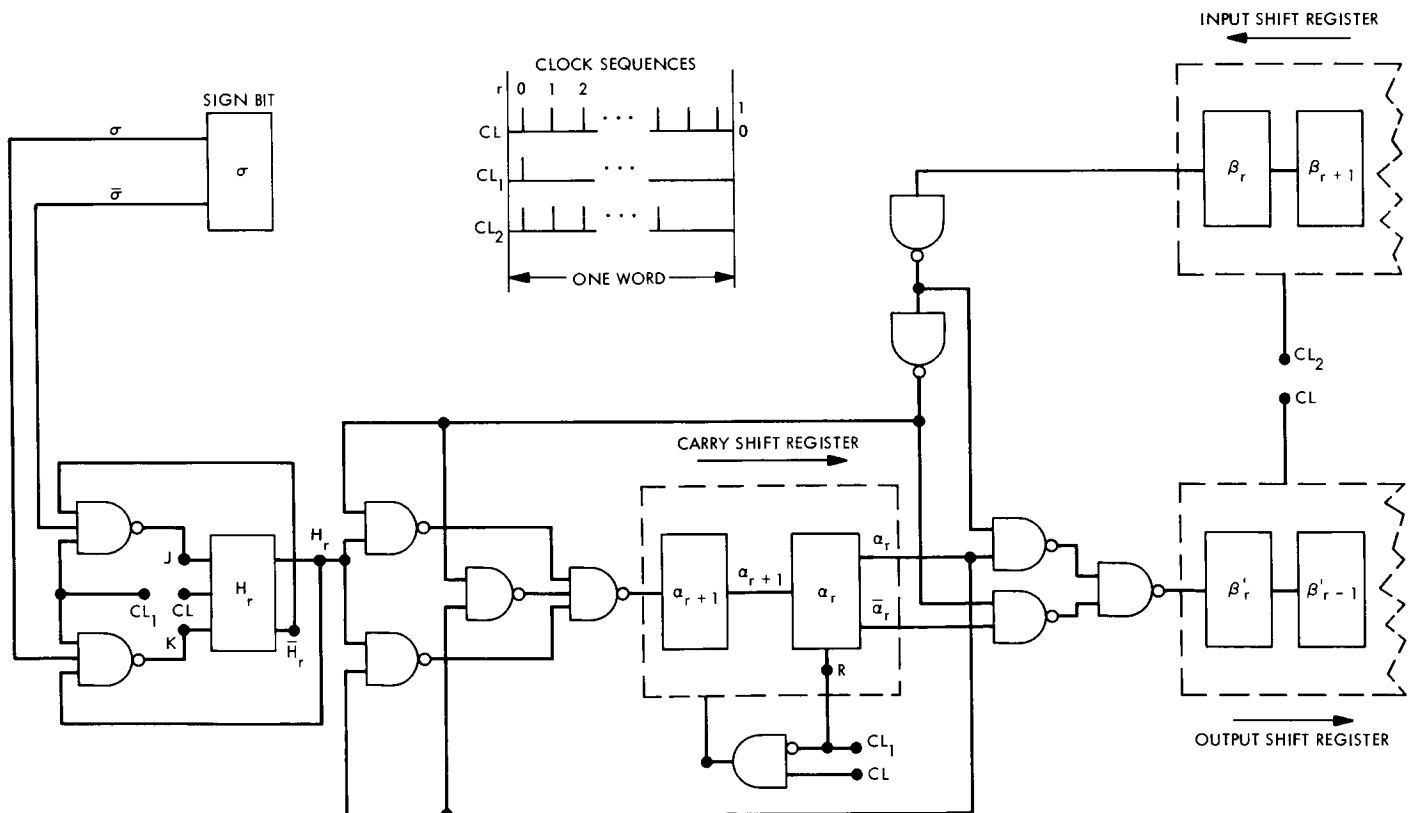


Fig. 6. Serial converter from sign-magnitude binary to base  $-2$

number will always have two bits more than the input magnitude.

The overall conversion, as shown in Fig. 6, seems quite simple. This suggests that the negative radix representation is a promising possibility to be considered in the design of various types of computers, special-purpose and otherwise.

## 5. The Required Number of Digits

In *Subsection 2*, we have seen an example of a number whose negative radix representation calls for more digits than the corresponding positive representation ( $3749 = 17869_{-10}$ ). Our purpose is to analyze this aspect of negative radix representation in some detail.

Let us examine a system that uses  $D$  digits to represent numbers in the base  $q < 0$ . The largest representable integer is the positive number  $P$  whose representation is

$$(|q| - 1, \dots, 0, |q| - 1, 0, |q| - 1)$$

Its value is given by

$$P = (|q| - 1) \frac{q^{2[(D+1)/2]} - 1}{q^2 - 1} = \frac{q^{2[(D+1)/2]} - 1}{|q| + 1} \quad (20)$$

Similarly, the smallest integer ( $N$ ) is

$$(|q| - 1, \dots, |q| - 1, 0, |q| - 1, 0)$$

with the corresponding value

$$N = q(|q| - 1) \frac{q^{2[D/2]} - 1}{q^2 - 1} = q \frac{q^{2[D/2]} - 1}{|q| + 1} \quad (21)$$

The number of integers contained in the closed interval defined this way is  $P - N + 1$ . Applying Eqs. (20) and (21) shows this expression to be  $|q|^D$ . This, however, is the number of different configurations of the  $D$  digits. We conclude, therefore, that  $D$  digits span all the integers from  $N$  to  $P$ . Note, however, that the closed interval  $(N, P)$  is quite asymmetrical. Specifically,

$$\frac{P}{|N|} = \begin{cases} \frac{1}{|q|} & (D \text{ even}) \\ \frac{1}{|q|} \cdot \frac{|q|^{D+1} - 1}{|q|^{D-1} - 1} \approx |q| & (D \text{ odd}) \end{cases} \quad (22)$$

Using the same  $D$  digits in a signless positive radix representation would accommodate the  $|q|^D$  integers spanning the interval  $(0, |q|^D - 1)$ . This leads to the following formulation of the effect of a negative radix on the interval of representation:

Changing the radix from  $|q|$  to  $-|q|$  shifts the interval of representable integers left into the negative region. The number of positive and negative integers in the interval are quite different. For even  $D$ , the negative integers are more numerous by a factor  $|q|$ . For odd  $D$ , the positive integers are more numerous by a factor  $\approx |q|$ .

A simple example will illustrate these statements. Assume

$$q = -10$$

$$D = 3$$

The largest number is

$$909_{-10} = 909$$

while the smallest is

$$90_{-10} = -90$$

Thus, the interval spanned is  $(-90, 909)$ . Similarly, for  $D = 4$ , we get the interval  $(-9090, 909)$ .

The corresponding intervals in signless positive radix representations are  $(0, 999)$ ,  $(0, 9999)$ , respectively.

We turn now to another important aspect of our problem illustrated by the above examples. Consider a given machine design that utilizes a representation consisting of a sign and two decimal digits. The interval here is  $(-99, 99)$ . Suppose now that we want to modify the design by adopting the radix  $-10$ . How many more digits do we need? From the above example, we see that  $D = 3$  is not quite sufficient and we have to adopt  $D = 4$ . Hence, elimination of the sign has forced upon us the addition of two more digits. This is a general property holding for any  $|q|$  and derivable from the application of the "borrow" and carry rules (Eqs. 9 and 10) to the algorithm discussed in *Subsection 3*.

We maintain, however, that though the above statement is mathematically correct, it would be wasteful to base design decisions on it. Usually, if a decimal machine is designed to cover the interval  $(-99, 99)$ , all it means is that the interval  $(-9, 9)$  is insufficient while the interval

(-999, 999) is unwarranted. Thus, unless the original interval was marginal to start with, the switch to negative radix may safely be accomplished by adding only one digit, getting the interval (-90, 909).

We turn now to the general case, that is, the original representation is in base  $|q|$  with  $D$  digits and a sign. The target representation is in base  $-|q|$  with  $D + 1$  digits.

In the original representation, the largest magnitude is  $|q|^D - 1$ . In the new representation, the magnitudes of the two interval boundaries ( $P, |N|$ ) are different. As in the example above, we are concerned with the smaller of the two. Examination of Eqs. (20) and (21) shows that the worst situation obtains for  $|N|$  combined with even  $D$ , where we have

$$|N| = \frac{|q|}{|q| + 1} (|q|^D - 1) \quad (23)$$

Hence, we conclude that the magnitude of the boundary closer to zero will, in the worst case, be  $|q|/(|q| + 1)$  times the original boundary magnitude. The multiplying factor for the other boundary magnitude is approximately  $q^2/(|q| + 1)$  (see Eq. 22).

Noting that

$$\frac{|q|}{|q| + 1} < 1; \quad \frac{q^2}{|q| + 1} > 1 \quad (|q| = 2, 3, \dots)$$

we see that the boundary closer to zero is closer than desired, while the other boundary is farther than needed. The overall effect is that performance evaluation based on the closer boundary will tend to be somewhat pessimistic.

The above analysis is particularly relevant to a binary machine ( $|q| = 2$ ). Here, the addition of an extra digit (bit) is compensated by the elimination of the sign bit. Hence, both the original and target systems have the same number of bits. Under these conditions, we find that the two multiplying factors are (approximately) 2/3; 4/3. The reduction here is more severe than in the example considered earlier since  $|q|$  is smaller.

We conclude that modifying a binary design to "negative binary" representation might call for an extra bit in addition to the reassigned sign bit. In this case, the

boundaries magnitudes will be approximately 4/3 and 8/3 times the original boundary magnitude.

## J. Propagation Studies: Time-of-Arrival

**Observations of Eleven Pulsars, P. E. Reichley,  
G. S. Downs, and G. A. Morris, Jr.**

### 1. Introduction

Several pulsars have been observed on a weekly basis at DSS 13 at the Goldstone DSCC<sup>10</sup>. Measurements were made of the pulse arrival times, using a cesium clock. The arrival times were then corrected to the barycenter of the solar system in a relativistic frame of reference. The resulting arrival times were analyzed by least-squares techniques to estimate the position and the period characteristics of each pulsar. We present here the results for CP 0328, AP 0823, CP 0950, CP 1133, AP 1237, PSR 1642 - 03, PSR 1749 - 28, PSR 1929 + 10, JP 1933, AP 2015, and PSR 2045 - 16.

### 2. Measurement Techniques

The observations were made at 2388 MHz (12.5 cm), using an antenna and receiver which have been described in Ref. 1. The signal was sampled digitally 5000 times per apparent pulse period. The rms level of the system noise was reduced by adding corresponding samples from 500 pulse periods on an on-line computer. The resulting 5000 samples were recorded on magnetic tape. This process was typically continued for an hour.

The time base was provided by a cesium clock running on UTC. This clock was compared monthly with clock 8 of the U.S. National Bureau of Standards (NBS). The maximum deviation of the cesium clock from NBS-8 was only 9  $\mu$ s over the span of the observations and was neglected. The data sampling was started by means of an arming device which utilized the 1-pulse/s output of the cesium clock. Thus, by arming the device, we were able to start the data sampling at a precise, predetermined time.

A set of predictions of the apparent frequency of pulsation was used to determine a mean pulse frequency to be used while taking data. The 1-MHz output of the cesium clock was used to drive a frequency synthesizer, the output of which was equal to  $10^6$  times the mean pulse frequency. The synthesizer output was used to determine the

<sup>10</sup>The authors thank the DSS 13 staff for their invaluable assistance with the observations.

sampling rate, as well as to maintain the phase of the mean frequency. The error in phase due to synthesizer round-off was less than 2 parts in  $10^6$  over an hour.

An average pulse shape was obtained for each pulsar by adding together several thousand pulses. It was found that the major component of each pulsar could be closely approximated by a triangular function. This triangular function was then correlated with the data to determine the phase of the received pulse train relative to the observations start time. The phase of the pulse train then yielded the time-of-arrival of the first pulse after the start time.

### 3. Data Analysis

A non-isotropic Schwarzschild line element was used as the geometrical model for the pulsar-receiving site geometry. The model contained variable parameters for period characteristics, position, distance, proper motion, and orbital motion of the pulsar. The orbit of the earth was computed from JPL Ephemeris DE-43. This ephemeris provides the position and velocity vectors of the planets in a non-isotropic Schwarzschild metric. The motion of the receiving site about the geocenter, the reduction to the epoch 1950.0, and the conversion from UTC to ET were computed by means of standard equations, using parameters developed at JPL (Ref. 2). A proper time to coordinate time correction (Ref. 3) was made for each time-of-arrival measurement.

The internal uncertainty of the earth's position in DE-43 contributed errors that were well below the statistical errors of the parameters we solved for. The external uncertainty between the earth's position and the epoch 1950.0, about 0.1 second of arc, was a limiting factor in position determination.

Since the pulsars differ in average energy, some time-of-arrival measurements were noise-limited and others were resolution-limited. In the noise-limited cases, all the 500-pulse samples recorded during the hour were added before the correlation with a triangle was performed. The resulting time-of-arrival was assigned a statistical weight proportional to the square of the signal-to-noise ratio. In the resolution-limited cases, the correlation with a triangle was performed with all of the 500-pulse samples. A time-of-arrival was determined by averaging the individual time-of-arrivals weighted by the squares of their associated signal-to-noise ratios.

The geometrical model was used to reduce the time-of-arrival measurements to the barycenter of the solar

system, as well as to predict pulse arrival times at the barycenter. Using the differences between predicted and measured times-of-arrival, the variable parameters of the model were differentially corrected by least-squares techniques. Any driftage in the pulse phase due to an error in the setting of the frequency synthesizer was removed during the least-squares analysis. In most cases, this driftage was due only to motion of the receiving site during the time of observation and was generally much less than the sampling resolution.

The error curves produced by right ascension and declination are periodic over a year's time. This means that the model parameters will be most separated, i.e., least correlated after 1 yr. In solving for parameters from data that spanned less than a year, high correlations were reflected in the error bars. In no case could we discern effects due to distance or proper motion of the pulsar.

### 4. Position Determinations

The positions determined by the least-squares analysis are shown in Table 3. The standard deviation of the fit and number of days of data apply to all the parameters we have determined (Tables 3 and 4). The errors quoted in Table 3 are one standard deviation, or 0.1 second of arc, whichever is larger.

The positions are in good agreement with the best position determinations that have appeared in the literature, with the exception of the declinations of 0950 + 08 and 2045 - 16 and the right ascension of 1642 - 03. These determinations do not agree (i.e., the error bars do not overlap) with previously quoted positions (Refs. 4-6). Our value of the right ascension of 1642 - 03 does agree if we increase our error bars to two standard deviations. To obtain agreement in the declinations of 0950 + 08 and 2045 - 16, we would have to increase our error bars to several standard deviations. However, the position of 0950 + 08, as well as those of 0329 + 54, 1133 + 16, 1933 + 16, and 2016 + 28, do agree with our earlier results (Ref. 7).

### 5. Period Characteristics

The period of each pulsar was considered to be of the form

$$P(t) = P + \dot{P}(t - t_0) + \frac{\ddot{P}}{2}(t - t_0)^2$$

where  $t_0$  was an initial epoch and  $t - t_0$  was expressed in UTC seconds. Hence, the units are:  $P$ , sec;  $\dot{P}$ , sec/sec;  $\ddot{P}$ , sec/sec<sup>2</sup>.

Table 3. Least-squares position determinations

Pulsar	$\alpha$ (1950.0)	$\delta$ (1950.0)	$\sigma$ of fit, $\mu$ s	Data span, days
0329 + 54	03 <sup>h</sup> 29 <sup>m</sup> 11 <sup>s</sup> .080 $\pm$ .007	54°24'38".3 $\pm$ ".1	487	415
0823 + 26	08 <sup>h</sup> 23 <sup>m</sup> 50 <sup>s</sup> .52 $\pm$ .02	26°47'18".1 $\pm$ ".8	138	256
0950 + 08	09 <sup>h</sup> 50 <sup>m</sup> 30 <sup>s</sup> .887 $\pm$ .007	8°09'56".0 $\pm$ ".3	210	415
1133 + 16	11 <sup>h</sup> 33 <sup>m</sup> 27 <sup>s</sup> .348 $\pm$ .007	16°07'29".6 $\pm$ ".2	223	366
1642 - 03	16 <sup>h</sup> 42 <sup>m</sup> 15 <sup>s</sup> $\pm$ 7 <sup>s</sup>	-3°12'36" $\pm$ 56"	1250	106
1749 - 28	17 <sup>h</sup> 49 <sup>m</sup> 49 <sup>s</sup> .03 $\pm$ .06	-28°06'07" $\pm$ 9"	144	168
1929 + 10	19 <sup>h</sup> 29 <sup>m</sup> 51 <sup>s</sup> .79 $\pm$ .14	10°53'04" $\pm$ 3"	116	119
1933 + 16	19 <sup>h</sup> 33 <sup>m</sup> 31 <sup>s</sup> .775 $\pm$ .007	16°09'58".8 $\pm$ ".1	161	307
2016 + 28	20 <sup>h</sup> 16 <sup>m</sup> 00 <sup>s</sup> .078 $\pm$ .007	28°30'31".0 $\pm$ ".1	288	415
2045 - 16	20 <sup>h</sup> 45 <sup>m</sup> 46 <sup>s</sup> .35 $\pm$ .07	-16°27'25" $\pm$ 4"	485	320

Table 4. Least-squares period characteristics

Pulsar	$P$ , sec $\pm$ msec	$\dot{P} \times 10^{15}$ , sec/sec	$ \ddot{P}  \times 10^{25}$ , sec/sec <sup>2</sup>	Epoch, UT
0329 + 54	0.71451855076 $\pm$ 0.02	2.061 $\pm$ 0.001	< 8	17:30 (9/5/68)
0823 + 26	0.53065958655 $\pm$ 0.05	1.664 $\pm$ 0.004	—	10:23 (2/12/69)
0950 + 08	0.253065032861 $\pm$ 0.005	0.2297 $\pm$ 0.0003	< 1.5	20:40 (9/5/68)
1133 + 16	1.18791101340 $\pm$ 0.04	3.741 $\pm$ 0.003	< 18	16:00 (10/25/68)
*1237 + 25	1.38244851 $\pm$ 170	—	—	00:00 (8/10/69)
1642 - 03	0.387688727 $\pm$ 16	6.2 $\pm$ 3.9	—	02:00 (7/12/69)
1749 - 28	0.5625533091 $\pm$ 0.3	8.23 $\pm$ 0.04	—	10:00 (5/11/69)
1929 + 10	0.2265170159 $\pm$ 0.2	1.16 $\pm$ 0.05	—	05:00 (6/29/69)
1933 + 16	0.35873517339 $\pm$ 0.02	6.003 $\pm$ 0.001	< 11	00:40 (12/23/68)
2016 + 28	0.55795338305 $\pm$ 0.01	0.1489 $\pm$ 0.0007	< 6	01:05 (9/6/68)
2045 - 16	1.9615663519 $\pm$ 0.4	10.98 $\pm$ 0.03	< 216	00:40 (12/10/68)

\*P based on  $\alpha$  (1950.0) = 12<sup>h</sup>37<sup>m</sup>17<sup>s</sup>  $\pm$  10<sup>s</sup> and  $\delta$  (1950.0) = 25°09'.5  $\pm$  4' (Ref. 8).

The results of our least-squares determination of the period characteristics ( $P, \dot{P}, \ddot{P}$ ) are shown in Table 4. The errors quoted are one standard deviation. The upper bound given for  $\ddot{P}$  is the one standard deviation error we obtained upon solving for  $\ddot{P}$ . The values we obtained for  $\ddot{P}$  were usually less than one standard deviation and appeared to be randomly distributed with normalized mean of zero and normalized variance of one. We attempted to solve for  $\ddot{P}$  only in cases where the span of the data was sufficient to insure a convergent solution. The solution for  $P$  of 1237 + 25 was made using the position given by Lang and Bosque (Ref. 8) and the error bar reflects the uncertainties in this position.

Our determinations of  $P$  are in good agreement with the best values that we are aware of, taking into account that  $P$  is in UTC seconds, referred to the barycenter, and given for a particular epoch. Our determination of  $P$  and  $\dot{P}$  for 0329 + 54, 0950 + 08, and 1133 + 16 agree well with the results of Davies, Hunt, and Smith (Ref. 9) and Cole (Refs. 10 and 11). These investigators used a technique similar to ours, although they did not solve for position. Also, our determination of  $P$  and  $\dot{P}$  for 0329 + 54, 0823 + 26, 0950 + 08, 1133 + 16, 1933 + 16, 2016 + 28, and 2045 - 16 agree well with our earlier results (Ref. 7).

The measurability of  $\ddot{P}$  can be calculated based on current pulsar theories (Ref. 12). Let  $\dot{P} = DP^\gamma$ , where  $D$  is constant and  $-3 \leq \gamma \leq 1$ , depending on the radiation process involved. If we take as a nominal value  $\gamma = -1$ , then  $\ddot{P} = -\dot{P}^2 P^{-1}$ . The term in the time-of-arrival model containing  $\ddot{P}$  is of the form

$$c(t) = \frac{(t - t_0)^3}{6} \frac{\ddot{P}}{P}$$

We can now calculate the time  $t - t_0$  required for  $\ddot{P}$  to be detectable. We require that  $c(t)$  be equal to the standard deviation of the least-squares fit (Table 3). This calculation yields times of 5 yr for 1749 - 28 and 1933 + 16 to 91 yr for 2015 + 28. It then seems that the detection of  $\ddot{P}$  for the pulsars presented here will be a lengthy process. Even an order of magnitude improvement in pulse arrival time measurements will reduce the time required for detection by only a factor of two.

## 6. Discussion

It has been pointed out by Davies, Hunt, and Smith (Ref. 9) that the error residual due to  $\dot{P}$  can be interpreted as orbital motion of the pulsar. They suggest that the parabolic error seen could be part of a sinusoidal error due to

orbital motion. Although this is possible, it becomes less probable each time a new source is investigated. All values of  $\dot{P}$  determined to date are positive while orbital motion should occasionally contribute an effect that would yield a negative  $\dot{P}$ .

There will be contributions to  $P$  and  $\dot{P}$  due to the changing distance between the pulsar and the solar system barycenter. The relative velocity will contribute a constant term to  $P$  and the relative acceleration will contribute a constant term to  $\dot{P}$ . If we assume a simple model of the galaxy (Ref. 13), the maximum contribution to  $P$  is  $\sim 2 \times 10^{-5}P$  and the maximum contribution to  $\dot{P}$  is  $\sim 10^{-17}P$  for the group of pulsars we have considered. These effects will be slowly changing and will probably never be detected.

It has been suggested that random variations in the length of the day by as much as 0.5 ms can affect our determination of  $\dot{P}$  (Ref. 14). Their analysis describes how accurately we may measure  $P$  but does not, however, apply to our measurements of  $\dot{P}$ . Let us examine the effect of a 0.5-ms error in the length of the day. Since the equatorial velocity of the earth is 0.456 km/s (Ref. 15), the maximum error in the assumed position of the receiving site is  $2.3 \times 10^{-4}$  km. This gives a pulse arrival time error  $\leq 7.7 \times 10^{-10}$  s. If the error persists for 1 yr, then the accumulated error in the pulse arrival time is 0.28  $\mu$ s. None of our pulse arrival time measurements are this accurate, and we conclude that such changes in the length of the day have remained undetected. Note that this result can be obtained by considering Maran and Öggleman's maximum value of  $(\Delta P/P)_{\text{random}} = 8.8 \times 10^{-15}$ . The error in the measured time-of-arrival after time  $T$  is  $n\Delta P$ , where  $n \approx T/P$ . In 1 yr, the accumulated error,  $\Delta P \times T/P$ , is 0.28  $\mu$ s.

The value of  $\dot{P}$  has been measured for sixteen (Refs. 9-11, 16, 17) pulsars and are listed in Table 5. The relationship between  $\dot{P}$  and  $P$  is presented in Fig. 7. Considering only the pulsars for which the period is greater than 0.1 s, there is no clear correlation between  $\dot{P}$  and  $P$ . The simple relationship of  $\dot{P} \propto P^{-1}$  suggested by the two fast pulsars 0531 + 21 and 0833 - 45 may become quite complicated at some critical rotation rate. It is known, for example, that variations in polarization with pulse phase are more complicated in the slower pulsars (Ref. 18). If we accept a model in which  $\dot{P} \approx DP^\gamma$ , then the slower pulsars ( $P > 0.1$  s) exhibit a range of values for  $D$  (Refs. 11 and 19). The parameter  $D$  is a function of the physical parameters of the model, while  $\gamma$  depends on the details of the energy loss. In particular, Ostriker and Gunn (Ref. 20) and

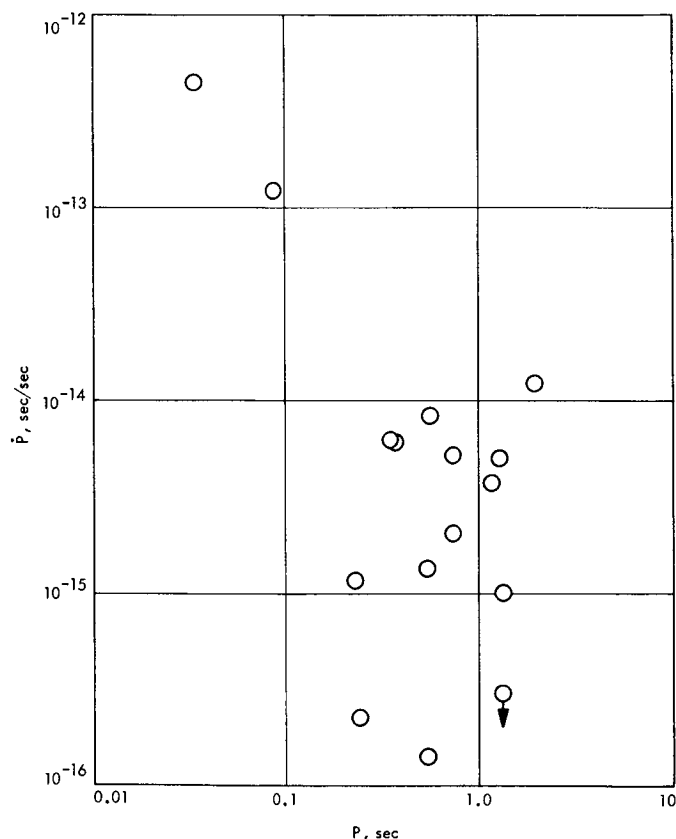
**Table 5. Surface magnetic fields and ages**

Pulsar	$P$ , sec	$\dot{P} \times 10^{15}$ , sec/sec	$B_s \div 10^{12}$ , gauss	Age $\div 10^6$ , yr
0329 + 54	0.714	2.06	0.76	5.6
0531 + 21	0.033	423.0	2.36	0.00125
0808 + 74	1.292	<0.3	<0.38	>69.0
0823 + 26	0.531	1.66	0.54	6.1
0833 - 45	0.089	125.0	2.1	0.0114
0834 + 06	1.274	5.0	1.6	4.05
0950 + 08	0.253	0.229	0.15	17.3
1133 + 16	1.188	3.74	1.34	5.0
1508 + 55	0.739	5.3	1.24	2.2
1642 - 03	0.387	6.2	0.98	1.0
1749 - 28	0.562	8.23	1.36	1.1
1919 + 21	1.337	1.0	0.72	21.0
1929 + 10	0.227	1.16	0.33	3.0
1933 + 16	0.358	6.00	0.29	0.95
2016 + 28	0.557	0.149	0.18	59.0
2045 - 16	1.961	10.9	2.92	2.8

Goldreich and Julian (Ref. 21) arrive at a simple relationship between  $P$ ,  $\dot{P}$ , and  $B_s$ , the surface value of the neutron star's magnetic field. The values of magnetic field presented in Table 5 are calculated from  $B_s = 2.2 \times 10^{19} (P \dot{P})^{1/2}$  gauss. The tabulated ages ( $P/2\dot{P}$ ) were calculated assuming that gravitational radiation is unimportant and that the magnetic field does not decay. The calculated surface strength of the magnetic field is confined to the remarkably narrow range of  $2 \times 10^{11}$  to  $3 \times 10^{12}$  gauss. This was noted by Ostriker and Gunn (Ref. 20), using a smaller sample of data. There is no clear correlation of magnetic field strength with pulse period. There is a suggestion of a correlation between magnetic field strength and age. Once again, this correlation disappears if 0531 + 21 and 0833 - 45 are omitted. Indeed, a recent calculation (Ref. 22) of the conductivity of a typical neutron star states that the decay time of the magnetic field is nearly  $10^{13}$  yr, a time  $10^8$  times longer than the estimated age of the universe.

#### References

1. Downs, G. S., Morris, G. A., and Reichley, P. E., *Nature*, Vol. 222, p. 1257, 1969.
2. Melbourne, W. G., et al., *Constants and Related Information for Astrodynamical Calculations*, 1968, Technical Report 32-1306. Jet Propulsion Laboratory, Pasadena, Calif., Jul. 15, 1968.
3. Clemence, G. M., and Szebehely, V., *Astron. J.*, Vol. 72, p. 1324, 1967.
4. Turtle, A. J., and Vaughan, A. E., *Nature*, Vol. 219, p. 689, 1968.
5. Turtle, A. J., and Vaughan, A. E., *Nature*, Vol. 219, p. 845, 1968.
6. Vaughan, A. E., Large, M. I., and Wielebinski, R., *Nature*, Vol. 222, p. 963, 1969.
7. Reichley, P. E., Downs, G. S., and Morris, G. A., Paper presented at USNC/URSI Meeting at Washington, D.C., Apr. 22, 1969.
8. Lang, K. R., and Bosque, B., *IAU Circular*, No. 2137, 1969.
9. Davies, J. G., Hunt, G. C., and Smith, F. G., *Nature*, Vol. 221, p. 27, 1969.
10. Cole, T. W., *Nature*, Vol. 221, p. 29, 1969.
11. Cole, T. W., *Nature*, Vol. 223, p. 487, 1969.
12. Boynton, P. E., et al., *Astrophys. J. (Letters)*, Vol. 157, p. L197, 1969.
13. Mihalas, D., and Routly, P. M., *Galactic Astronomy*, pp 127-138. W. H. Freeman and Co., San Francisco, Calif., 1968.
14. Maran, S. P., and Ögelman, H., *Nature*, Vol. 224, p. 349, 1969.
15. Allen, C. W., *Astrophysical Quantities*. Athlone Press, London, England, 1963.
16. Richards, D. W., and Comella, J. M., *Nature*, Vol. 222, p. 551, 1969.
17. Radhakrishnan, V., et al., *Nature*, Vol. 221, p. 443, 1969.



**Fig. 7.  $\dot{P}$  versus  $P$  for sixteen pulsars**

Subsection 6 discusses and compares the efficiencies of the various algorithms considered while Subsection 7 extends the discussion to the case where several sets of equations with a common  $L_m$  are to be solved.

## 2. Preliminaries

Both the Trench and the Levinson algorithms require that all principal minors of  $L_m$  be non-zero. This constraint carries over to the algorithm developed here and allows us to normalize  $L_m$  by dividing it by its main diagonal term. Applying the same division to  $d_m$  guarantees that  $s_m$  is not affected. In the following, we assume that this normalization process has been carried out so that  $L_m$  has units along its main diagonal.

We adopt here the following notation conventions:

Greek letters are used for scalars, capital letters for square matrices, and lower case letters for column matrices. Transposition is indicated by the symbol  $\sim$ . Thus,

$L_k$  is a  $k \times k$  matrix

$g_k$  is a  $k \times 1$  matrix

$\tilde{g}_k$  is a  $1 \times k$  matrix

$\lambda_k$  is a scalar

Another symbol which will appear quite often is the reversal symbol  $\wedge$ . Given, say, the column matrix  $g_k$ , we use  $\hat{g}_k$  to denote the reversed order version of  $g_k$ , that is,

$$(\hat{g}_k)_{i,1} = (g_k)_{k+1-i,1} \quad (2)$$

Finally, the letter  $m$  always denotes the number of equations in the system to be solved while the associated variable  $n$  always satisfies

$$m = n + 1 \quad (2a)$$

## 3. Trench Algorithm

Given the normalized Toeplitz matrix  $L_m$ , we consider its  $i$ th order upper-left submatrix  $L_i$  and denote the corresponding inverse  $B_i$ .

$$B_i = L_i^{-1} \quad (3)$$

Note that all  $L_i$  defined in this way are normalized Toeplitz matrices.

The derivation of the Trench algorithm starts as a straightforward application of the bordering method of matrix inversion. This is an iterative process in which  $B_{i+1}$  is computed as a function of  $B_i$ . Starting with the elementary  $B_2$ , this eventually leads to the desired  $B_m$ . The application of this scheme to Toeplitz matrices is drastically simplified by the fact that the first row and column of the inverse of a Toeplitz matrix,  $B_i$ , uniquely determine all remaining  $B_i$  elements (Ref. 1). Furthermore, these other elements of  $B_i$  are not needed when we turn to compute the first row and column of  $B_{i+1}$ . Thus, instead of an iteration process involving the square matrices  $B_i$ , we have to deal with an iteration process involving column matrices only.

In view of this, we denote

$$B_{i+1} = \frac{1}{\lambda_i} \begin{bmatrix} 1 & \tilde{e}_i \\ g_i & M_i \end{bmatrix} \quad (4)$$

The first phase of the Trench algorithm is an iterative scheme in which  $\lambda_n, e_n, g_n$  are recursively determined from the initial trivial values  $\lambda_1, e_1, g_1$ . In the second and final phase of the algorithm, the elements of  $M_n$  are computed from  $\lambda_n, e_n, g_n$ . This yields  $B_m$ .

The following is a summary of the algorithm.

### TRENCH ALGORITHM (NON-HERMITIAN CASE).

#### Problem formulation.

$$L_{n+1} = \begin{bmatrix} 1 & \tilde{a}_n \\ r_n & L_n \end{bmatrix}$$

$$\tilde{a}_i = [\rho_{-1} \rho_{-2} \cdots \rho_{-i}] \quad (1 \leq i \leq n)$$

$$\tilde{r}_i = [\rho_1 \rho_2 \cdots \rho_i] \quad (1 \leq i \leq n)$$

$$B_{n+1} = L_{n+1}^{-1} = ?$$

#### Phase 1.

##### Initial values for recursion.

$$e_1 = -\rho_{-1}$$

$$g_1 = -\rho_1$$

$$\lambda_1 = 1 - \rho_{-1} \rho_1$$

Recursion of  $\lambda, g, e$  ( $1 \leq i < n$ ).

$$\eta_i = -(\rho_{-(i+1)} + \tilde{e}_i \hat{a}_i) \quad \gamma_i = -(\rho_{i+1} + \tilde{r}_i \hat{g}_i)$$

$$e_{i+1} = \begin{bmatrix} e_i + \frac{\eta_i}{\lambda_i} \hat{g}_i \\ \frac{\eta_i}{\lambda_i} \end{bmatrix} \quad \hat{g}_{i+1} = \begin{bmatrix} \frac{\gamma_i}{\lambda_i} \\ \hat{g}_i + \frac{\gamma_i}{\lambda_i} e_i \end{bmatrix}$$

$$\lambda_{i+1} = \lambda_i - \frac{\eta_i \gamma_i}{\lambda_i}$$

**Phase 2 (construction of  $B_{n+1}$ ).** The first row and column are obtained from

$$\lambda_n B_{n+1} = \begin{bmatrix} 1 & \tilde{e}_n \\ g_n & M_n \end{bmatrix}$$

All other terms are obtained from

$$\lambda_n (B_{n+1})_{i+1, j+1} = \lambda_n (B_{n+1})_{ij} + (g_n \tilde{e}_n - \hat{e}_n \tilde{g}_n)_{ij} \\ (B_{n+1})_{ij} = (B_{n+1})_{n+2-j, n+2-i}$$

The last equation of Phase 2 is a statement of a general property, namely, the inverse of a Toeplitz matrix is persymmetric, that is, it has symmetry about the cross-diagonal extending from the upper right corner to the lower left corner (see Ref. 1 for proof).

#### 4. Generalized Levinson Algorithm

Consider the system of  $m (= n + 1)$  linear equations

$$L_m s_m = d_m \quad (5)$$

where

$$\tilde{d}_m = [\delta_1 \delta_2 \cdots \delta_m]$$

and  $L_m$  is an  $m$ th order normalized Toeplitz matrix defined in terms of  $a_n, r_n$  as in Subsection 3. The approach here is similar to the bordering method underlying the Trench algorithm. Thus, denoting

$$\tilde{d}_i = [\delta_1 \delta_2 \cdots \delta_i] \quad (1 \leq i \leq m) \quad (6)$$

we consider the systems of equations

$$L_i s_i = d_i \quad (1 \leq i \leq m) \quad (7)$$

Starting with the trivial case

$$s_1 = d_1 = \delta_1 \quad (8)$$

we seek to establish a recursion that would allow us to compute  $s_{i+1}$  in terms of  $s_i$ . This would then eventually lead to the desired solution  $s_m$ .

In general, we may formulate

$$s_{i+1} = \begin{bmatrix} s_i \\ 0 \end{bmatrix} + \psi_i \begin{bmatrix} f_i \\ 1 \end{bmatrix} \quad (9)$$

where  $\psi_i, f_i$  are as yet unknown.

We proceed now to determine  $\psi_i, f_i$ , starting with the formulation of Eq. (7) for the index  $i + 1$ ,

$$d_{i+1} = \begin{bmatrix} d_i \\ \delta_{i+1} \end{bmatrix} = L_{i+1} s_{i+1} \quad (10)$$

Substituting Eq. (9) in Eq. (10), we get

$$\psi_i L_{i+1} \begin{bmatrix} f_i \\ 1 \end{bmatrix} = \begin{bmatrix} d_i \\ \delta_{i+1} \end{bmatrix} - L_{i+1} \begin{bmatrix} s_i \\ 0 \end{bmatrix} \quad (11)$$

We recall now that the Toeplitz structure of  $L_m$  implies that

$$L_{i+1} = \begin{bmatrix} 1 & \tilde{a}_i \\ r_i & L_i \end{bmatrix} = \begin{bmatrix} L_i & \hat{a}_i \\ \tilde{r}_i & 1 \end{bmatrix} \quad (12)$$

Substituting this on the right of Eq. (11), we get

$$\psi_i L_{i+1} \begin{bmatrix} f_i \\ 1 \end{bmatrix} = \begin{bmatrix} d_i \\ \delta_{i+1} \end{bmatrix} - \begin{bmatrix} L_i & \hat{a}_i \\ \tilde{r}_i & 1 \end{bmatrix} \begin{bmatrix} s_i \\ 0 \end{bmatrix} = \begin{bmatrix} 0_i \\ \theta_i \end{bmatrix} \quad (13)$$

where  $0_i$  is the  $i$ th order zero column matrix and

$$\theta_i = \delta_{i+1} - \tilde{r}_i s_i = \delta_{i+1} - \tilde{s}_i \hat{r}_i \quad (14)$$

Hence,

$$\psi_i \begin{bmatrix} f_i \\ 1 \end{bmatrix} = B_{i+1} \begin{bmatrix} 0_i \\ \theta_i \end{bmatrix} \quad (15)$$

The appearance of  $0_i$  on the right means that only the last column of  $B_{i+1}$  is needed. We have seen (Eq. 4) that  $B_{i+1}$  is expressible in terms of the Trench algorithm parameters  $\lambda_i, e_i, g_i$ :

$$B_{i+1} = \frac{1}{\lambda_i} \begin{bmatrix} 1 & \tilde{e}_i \\ g_i & M_i \end{bmatrix}$$

Invoking now the persymmetry of  $B_{i+1}$  (see end of *Subsection 3*), we see that an equally valid alternative form is

$$B_{i+1} = \frac{1}{\lambda_i} \begin{bmatrix} P_i & \hat{e}_i \\ \tilde{g}_i & 1 \end{bmatrix} \quad (16)$$

where  $P_i$  is some function of  $M_i, e_i, g_i$ . Substituting this in Eq. (15), we get

$$\psi_i \begin{bmatrix} f_i \\ 1 \end{bmatrix} = \frac{1}{\lambda_i} \begin{bmatrix} P_i & \hat{e}_i \\ \tilde{g}_i & 1 \end{bmatrix} \begin{bmatrix} 0_i \\ \theta_i \end{bmatrix} = \frac{\theta_i}{\lambda_i} \begin{bmatrix} \hat{e}_i \\ 1 \end{bmatrix}$$

Hence,

$$\psi_i = \frac{\theta_i}{\lambda_i} \quad (17)$$

$$f_i = \hat{e}_i \quad (18)$$

and the recursion of  $s_i$  takes the following form:

$$s_{i+1} = \begin{bmatrix} s_i + \frac{\theta_i}{\lambda_i} \hat{e}_i \\ \frac{\theta_i}{\lambda_i} \end{bmatrix} \quad (19)$$

The appearance of  $\hat{e}_i$  in Eq. (19) means that, in addition to the iteration of  $s_i$ , we have to apply the Trench algorithm recursion formulae to compute  $\hat{e}_i$ ; and since this calls for  $g_i$ , we see that the algorithm developed here must incorporate all of Phase 1 of the Trench algorithm.

The following is a summary of the algorithm.

**GENERALIZED LEVINSON ALGORITHM (NON-HERMITIAN CASE).**

*Problem formulation.*

$$L_{n+1} s_{n+1} = d_{n+1}$$

$$L_{n+1} = \begin{bmatrix} 1 & \tilde{a}_n \\ r_n & L_n \end{bmatrix}$$

$$\tilde{a}_i = [\rho_{-1} \rho_{-2} \cdots \rho_{-i}] \quad (1 \leq i \leq n)$$

$$\tilde{r}_i = [\rho_1 \rho_2 \cdots \rho_i] \quad (1 \leq i \leq n)$$

$$\tilde{d}_{n+1} = [\delta_1 \delta_2 \cdots \delta_{n+1}]$$

$$s_{n+1} = ?$$

*Initial values for recursion.*

$$s_1 = \delta_1$$

$$e_1 = -\rho_{-1}$$

$$g_1 = -\rho_1$$

$$\lambda_1 = 1 - \rho_{-1}\rho_1$$

*Recursion of  $s, \hat{e}, g, \lambda$  ( $i = 1, 2, \cdots$ ).*

$$\theta_i = \delta_{i+1} - \tilde{s}_i \hat{r}_i$$

$$\eta_i = -\rho_{-(i+1)} - \tilde{a}_i \hat{e}_i$$

$$\gamma_i = -\rho_{i+1} - \tilde{g}_i \hat{r}_i$$

$$s_{i+1} = \begin{bmatrix} s_i + \frac{\theta_i}{\lambda_i} \hat{e}_i \\ \frac{\theta_i}{\lambda_i} \end{bmatrix}$$

$$\hat{e}_{i+1} = \begin{bmatrix} \frac{\eta_i}{\lambda_i} \\ \hat{e}_i + \frac{\eta_i}{\lambda_i} g_i \end{bmatrix}$$

$$g_{i+1} = \begin{bmatrix} g_i + \frac{\gamma_i}{\lambda_i} \hat{e}_i \\ \frac{\gamma_i}{\lambda_i} \end{bmatrix}$$

$$\lambda_{i+1} = \lambda_i - \frac{\eta_i \gamma_i}{\lambda_i}$$

**Last computed values.**

$$\theta_n, \eta_{n-1}, \gamma_{n-1}$$

$$s_{n+1}, \hat{e}_n, g_{n-1}$$

$$\lambda_n$$

**EXAMPLE.** Let the given system of equations be

$$\begin{bmatrix} 1 & -3 & 1 & 2 \\ -2 & 1 & -3 & 1 \\ -1 & -2 & 1 & -3 \\ 1 & -1 & -2 & 1 \end{bmatrix} \begin{bmatrix} \\ \\ s_4 \\ \end{bmatrix} = \begin{bmatrix} 5 \\ -6 \\ -10 \\ 2 \end{bmatrix}$$

The computer implementation of this algorithm will not be discussed since it is very similar to the implementation of the Trench algorithm described in Ref. 1.

Obviously,

$$n = 3$$

$$\tilde{a}_3 = [-3 \quad 1 \quad 2]$$

$$\tilde{r}_3 = [-2 \quad -1 \quad 1]$$

$$\tilde{d}_4 = [5 \quad -6 \quad -10 \quad 2]$$

We conclude this subsection with a simple numerical example to illustrate the algorithm.

**Initial values.**

$$s_1 = 5$$

$$e_1 = 3$$

$$g_1 = 2$$

$$\lambda_1 = 1 - 3 \cdot 2 = -5$$

**Recursion.**

$$\theta_1 = -6 + 5 \cdot 2 = 4$$

$$\eta_1 = -1 + 3 \cdot 3 = 8$$

$$\gamma_1 = 1 + 2 \cdot 2 = 5$$

$$s_2 = \begin{bmatrix} 5 - \frac{4}{5} \cdot 3 \\ -\frac{4}{5} \end{bmatrix} = \begin{bmatrix} \frac{13}{5} \\ -\frac{4}{5} \end{bmatrix}$$

$$\hat{e}_2 = \begin{bmatrix} -\frac{8}{5} \\ 3 - \frac{8}{5} \cdot 2 \end{bmatrix} = \begin{bmatrix} -\frac{8}{5} \\ -\frac{1}{5} \end{bmatrix}$$

$$g_2 = \begin{bmatrix} 2 - 1 \cdot 3 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$\lambda_2 = -5 + \frac{8 \cdot 5}{5} = 3$$

$$\theta_2 = -10 - \left( -\frac{13}{5} \cdot 1 + \frac{4}{5} \cdot 2 \right) = -9$$

$$\eta_2 = -2 - \left( 3 \cdot \frac{8}{5} - 1 \cdot \frac{1}{5} \right) = -\frac{33}{5}$$

$$\gamma_2 = -1 - (1 \cdot 1 + 1 \cdot 2) = -4$$

$$s_3 = \begin{bmatrix} \frac{13}{5} + 3 \cdot \frac{8}{4} \\ -\frac{4}{5} + 3 \cdot \frac{1}{5} \\ -3 \end{bmatrix} = \begin{bmatrix} \frac{37}{5} \\ -\frac{1}{5} \\ -3 \end{bmatrix}$$

$$\hat{e}_3 = \begin{bmatrix} -\frac{11}{5} \\ -\frac{8}{5} + \frac{11}{5} \cdot 1 \\ -\frac{1}{5} + \frac{11}{5} \cdot 1 \end{bmatrix} = \begin{bmatrix} -\frac{11}{5} \\ \frac{3}{5} \\ 2 \end{bmatrix}$$

$$\lambda_3 = 3 - \frac{\frac{33}{5} \cdot 4}{3} = -\frac{29}{5}$$

$$\theta_3 = 2 - \left( \frac{37}{5} \cdot 1 + \frac{1}{5} \cdot 1 + 3 \cdot 2 \right) = -\frac{58}{5}$$

$$s_4 = \begin{bmatrix} \frac{37}{5} - 2 \cdot \frac{11}{5} \\ -\frac{1}{5} + 2 \cdot \frac{3}{5} \\ -3 + 2 \cdot 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \\ 2 \end{bmatrix} \quad \leftarrow \text{the solution}$$

Direct substitution verifies the result.

$$\begin{bmatrix} 1 & -3 & 1 & 2 \\ -2 & 1 & -3 & 1 \\ -1 & -2 & 1 & -3 \\ 1 & -1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ -6 \\ -10 \\ 2 \end{bmatrix}$$

### 5. Hermitian Case

When  $L_{n+1}$  is Hermitian, the above algorithm will be simplified. We note that a Hermitian  $L_{n+1}$  is characterized by

$$a_n = r_n^* \quad (*\text{denotes complex conjugation}) \quad (20)$$

This leads to

$$e_i = g_i^* \quad (1 \leq i \leq n) \quad (21)$$

Substituting these in the formulae of *Subsection 4*, we get the following summary for this important special case.

### MODIFIED LEVINSON ALGORITHM.

#### Problem formulation.

$$\begin{aligned} L_{n+1} s_{n+1} &= d_{n+1} \\ L_{n+1} &= \begin{bmatrix} 1 & \tilde{r}_n^* \\ r_n & L_n \end{bmatrix} \\ \tilde{r}_i &= [\rho_1 \rho_2 \cdots \rho_i] \quad (1 \leq i \leq n) \\ \tilde{d}_{n+1} &= [\delta_1 \delta_2 \cdots \delta_{n+1}] \\ s_{n+1} &= ? \end{aligned}$$

#### Initial values for recursion.

$$\begin{aligned} s_1 &= \delta_1 & g_1 &= -\rho_1 \\ \lambda_1 &= 1 - |\rho_1|^2 \end{aligned}$$

### Recursion of $s, g, \lambda$ ( $i = 1, 2, \dots$ ).

$$\begin{aligned} \theta_i &= \delta_{i+1} - \tilde{s}_i \hat{r}_i & \gamma_i &= -(\rho_{i+1} + \tilde{r}_i \hat{g}_i) \\ s_{i+1} &= \begin{bmatrix} s_i + \frac{\theta_i}{\lambda_i} \hat{g}_i^* \\ \frac{\theta_i}{\lambda_i} \end{bmatrix} & \hat{g}_{i+1} &= \begin{bmatrix} \frac{\gamma_i}{\lambda_i} \\ \hat{g}_i + \frac{\gamma_i}{\lambda_i} \hat{g}_i^* \end{bmatrix} \\ \lambda_{i+1} &= \lambda_i - \frac{|\gamma_i|^2}{\lambda_i} \end{aligned}$$

#### Last computed values.

$$\begin{aligned} \theta_n, \gamma_{n-1} \\ s_{n+1}, \hat{g}_n \\ \lambda_n \end{aligned}$$

When  $L_{n+1}, d_{n+1}$  are real, the above algorithm assumes almost exactly the form of the original Levinson algorithm. There is, however, one important difference relating to the computation of  $\lambda_i$ .

Following Trench, we compute  $\lambda_i$  by the recursion

$$\lambda_{i+1} = \lambda_i - \frac{|\gamma_i|^2}{\lambda_i} \quad (22)$$

Levinson, on the other hand, uses the formula

$$\lambda_i = 1 + \tilde{r}_i g_i \quad (23)$$

Let us first verify this formula. The simplest derivation follows from

$$L_{i+1} B_{i+1} = I_{i+1}$$

where  $I_k$  is the  $k$ th order identity matrix. We express  $L_{i+1}$  in one of the forms shown in Eq. (12). Using Eq. (4) for  $B_{i+1}$ , we get

$$\begin{bmatrix} 1 & \tilde{a}_i \\ r_i & L_i \end{bmatrix} \frac{1}{\lambda_i} \begin{bmatrix} 1 & \tilde{e}_i \\ g_i & M_i \end{bmatrix} = \begin{bmatrix} 1 & \tilde{0}_i \\ 0_i & I_i \end{bmatrix}$$

The (1, 1) element of  $I_{i+1}$  yields

$$\lambda_i = 1 + \tilde{a}_i g_i \quad (24)$$

Alternatively, using Eq. (16) for  $B_{i+1}$ , we have

$$\begin{bmatrix} L_i & \hat{a}_i \\ \tilde{r}_i & 1 \end{bmatrix} \frac{1}{\lambda_i} \begin{bmatrix} P_i & \hat{e}_i \\ \tilde{g}_i & 1 \end{bmatrix} = \begin{bmatrix} I_i & 0_i \\ \tilde{0}_i & 1 \end{bmatrix}$$

The  $(i+1, i+1)$  element of  $I_{i+1}$  now yields

$$\lambda_i = 1 + \tilde{r}_i \hat{e}_i = 1 + \tilde{r}_i e_i \quad (25)$$

When  $L_m$  is real symmetric, both Eqs. (24) and (25) reduce to Eq. (23), which is thus verified.

We turn now to a comparison of Eqs. (22) and (23) from the point of view of computation efficiency. We consider first the number of computations involved in the adoption of the Recursion (22). Since  $\gamma_i$  is required for other parts of the algorithm, it is obvious that Eq. (22) prescribes one "operation" plus one division for each new  $\lambda_i$ . Hence, in solving a set of  $m$  equations, we have to invest  $(m-1)$  "operations" plus  $(m-2)$  divisions in computing all required  $\lambda_i$ 's. (There is no division in computing  $\lambda_1$ ; hence, the  $(m-2)$  term.)

Equation (23), on the other hand, requires  $i$  "operations" for each  $\lambda_i$ , leading to a total investment of  $m^2/2 - m/2$  "operations." For large  $m$ , the  $m^2$  term overshadows all others. Hence, we conclude that our modification of the Levinson algorithm trims the number of "operations" by  $m^2/2$ .

## 6. Comparison and Evaluation

As a first step in evaluating the various algorithms, we determine the number of computations required by each one as a function of the number of equations  $m$ . The results appear in Table 6 for the nonsymmetric case and Table 7 for the symmetric one.

The expressions appearing in Tables 6 and 7 essentially follow in a straightforward manner from the algorithms' specifications. We review briefly some minor details which have to be known if verification is attempted.

First, it should be mentioned that, in the interest of a more realistic account, we have assumed the given Toeplitz matrix to be un-normalized. Tables 6 and 7, therefore, account for the extra computations involved in the normalization.

In normalizing, we multiply by the reciprocal of the normalizing constant rather than divide by the normalizing constant to minimize the number of divisions. Furthermore, in this process we compute  $-a_n, -r_n$  rather than  $a_n, r_n$  since this results in savings by eliminating some multiplications by  $(-1)$ .

In the case of the Levinson algorithms, the normalization of  $L_m$  is accompanied by an equivalent treatment of  $d_m$ , guaranteeing the correct  $s_m$ .

A different strategy is indicated for the Trench algorithms. In this case, the structure of Phase 2 makes it more profitable to compute  $\lambda_n s_m$  first. Therefore, we forego modification of  $d_m$  and compensate for the normalization of  $L_m$  in the transformation from  $(\lambda_n s_m)$  to  $s_m$ , saving about  $m$  multiplications.

Examination of Tables 6 and 7 reveals that the numbers of required additions and multiplications are quadratics in the number of equations  $m$  while the numbers of divisions are linear functions of  $m$ .

**Table 6. Number of computations to solve  $L_m s_m = d_m$  ( $L_m$  non-symmetric)**

Algorithm	Additions	Multiplications	Divisions
Trench	$4m^2 - 9m + 5$	$4m^2 - 5m + 2$	$3m - 4$
Generalized Levinson	$3m^2 - 7m + 5$	$3m^2 - 4m + 4$	$4m - 7$

**Table 7. Number of computations to solve  $L_m s_m = d_m$  ( $L_m$  symmetric)**

Algorithm	Additions	Multiplications	Divisions
Trench	$[2.5m^2 - 4m + 1.5]$	$[2.5m^2 - m - 0.5]$	$2m - 2$
Levinson	$2.5m^2 - 4.5m + 2$	$2.5m^2 - 2.5m + 2$	$2m - 2$
Modified Levinson	$2m^2 - 3m + 1$	$2m^2 - m + 1$	$3m - 4$

When  $m$  is large, only the highest power terms are of interest. In this case then, the additions and multiplications appreciably outnumber the divisions. Furthermore, in the limit of large  $m$ , the number of additions equals the number of multiplications. The overall effect is that we are justified in this case in simply specifying the required computations in terms of "operations," as defined in *Subsection 1*. Thus, looking up the generalized Levinson algorithm in Table 6, we see that for large  $m$  the required computations are just  $3m^2$  "operations." The breakdown of this term is quite simple. Each of the iterations culminating in  $s_{n+1}, \hat{e}_n, g_{n-1}$ , requires  $m^2$  "operations." All other computations, as well as differences between the iterations of these three variables, affect only lower  $m$  powers.

In the special case where  $L_m$  is symmetric, we have to iterate only over  $s_i, g_i$ . Hence, the total reduces to  $2m^2$  "operations." The computation of  $\lambda_i$  does not show up here since it accounts for lower powers of  $m$  in the modified algorithm. In the original version, however, the computation of  $\lambda_i$  requires  $m^2/2$  "operations"; hence, the total of  $2.5m^2$  "operations" indicated in Table 7 for this case.

The composition of the dominant term is equally simple in the case of the Trench algorithm. Consider the non-symmetric case first. The iterations leading to  $e_n, \hat{g}_n$  account for  $2m^2$  "operations" (Phase 1). The construction of  $\lambda_n B_m$  from  $e_n, \hat{g}_n$  requires 2 "operations" for each evaluated term (see *Subsection 3*). However, since the inverse  $B_m$  is persymmetric, only  $m^2/2$  terms<sup>13</sup> have to be computed, leading to  $m^2$  "operations" for the actual inverse construction of Phase 2. Finally, multiplying the inverse by  $d_m$  adds another  $m^2$  term, bringing the total to  $4m^2$  "operations."

In the symmetric case,  $g_n$  contributes  $m^2$  "operations," while the computation of  $\lambda_n B_m$  contributes  $m^2/2$  "operations." (The inverse is now both symmetric and persymmetric so that only  $m^2/4$  terms<sup>14</sup> have to be evaluated.) Adding now the  $m^2$  term for multiplying by  $d_m$ , we get the indicated value of  $2.5m^2$  "operations."

Obviously, the generalized Levinson algorithm and its symmetric special case are superior to the other algorithms considered. In the non-symmetric case, it leads to a 25% reduction in computations when compared to the

best of the previously available methods. The reduction in the symmetric case is smaller but still significant, namely, 20%.

We close this subsection with a few words regarding storage requirements. We exclude from consideration the storage of input and output quantities such as  $a_n, r_n, s_m$  and turn our attention to the storage needed to accommodate intermediate results such as  $\hat{e}_i, g_i$ .

The required number of cells will be a polynomial in  $m$ , but a detailed analysis is unwarranted. We consider here only the dominant term of this polynomial.

With the above qualifications, we may say that all non-symmetric algorithms require  $2m$  memory cells while the symmetric algorithms require  $m$  cells.

Let us examine the justification of this statement for the generalized Levinson algorithm. We note first that once the elements of  $g_i$  have been used in an iteration, they are no longer needed; hence, the elements of  $g_{i+1}$  may be destructively overlayed on the elements of  $g_i$ . This means that a single array of size  $m$  will accommodate all  $g_i$ 's.<sup>15</sup>

Similar statements are valid for the  $\hat{e}_i$ 's and  $s_i$ 's. Realizing that the  $s_i$ 's can reside in the output array which will hold the final result  $s_m$ , we see that (for large  $m$ ) the only storage requirements are those for  $g_{n-1}, \hat{e}_n$ , namely,  $2m$  cells.

The same result holds for the Trench algorithm, provided we combine the construction of the inverse  $B_m$  in Phase 2 with its multiplication by  $d_m$  in a single subroutine to eliminate the storage of  $B_m$ .

The results for the symmetric algorithms are trivial extensions of the above.

## 7. Sets of Equations With a Common $L_m$

It should be recalled that our starting point was based on the implicit understanding that if several sets of equations with a common  $L_m$  are to be solved, then the Trench algorithm provides the proper efficient tool. It was only the case of the single set of equations which prompted the search for a better tool.

Having found this better tool in the form of the generalized Levinson algorithm, we wonder whether it might

<sup>13</sup>This is an approximation for large  $m$ .

<sup>14</sup>These are approximations for large  $m$ .

<sup>15</sup>The actual size is  $m-2$ , but we specify only the dominant term  $m$ .

not prove to be the better tool even for the case of  $k$  sets of equations with a common  $L_m$ .

Let us examine then the application of the generalized Levinson algorithm to this  $k$ -fold solution. We start with the non-symmetric case, noting that since all  $k$  sets share the same  $L_m$ , they must have the same  $\lambda_i, \hat{e}_i, g_i$ . Yet, if we adopt a straightforward approach and independently apply the algorithm to each of the  $k$  sets, we will be repeating the computation of these entities  $k$  times. Since the computation of  $\lambda_i, \hat{e}_i, g_i$  accounts for  $2m^2$  "operations" for the solution of one set, this course of action is very inefficient.

An obvious solution is to store all the  $\lambda_i$  scalars and the  $\hat{e}_i$  column matrices<sup>16</sup> ( $1 \leq i \leq n$ ) upon their first computation so that, while the solution of the first set would still require  $3m^2$  "operations," all succeeding sets would involve only the  $s_i$  iteration and would thus require  $m^2$  "operations" per set. The penalty is the drastically increased storage requirement. Thus, while the slow, inefficient (minimum storage) method requires  $2m$  memory cells, the faster (minimum time) method requires  $m^2/2$  cells.

The symmetric case presents a trivial extension of the above. Both cases are summarized in Table 8.

To complete the investigation, we have to apply now the same analysis to the Trench algorithm. Here, too, we have two variants. In the minimum time variation, the inverse  $B_m$  is computed in solving the first set and stored for use with the succeeding sets of equations. As we have already seen in Subsection 6, this requires  $m^2/2$  cells in the general case and  $m^2/4$  cells when  $L_m$  is symmetric.

In the minimum storage variant, only  $\lambda_n, e_n, \hat{g}_n$  are stored. This means that each successive set of equations will require  $2m^2$  "operations" in the non-symmetric case and  $1.5m^2$  "operations" in the symmetric case.

The summary of all four schemes in Table 8 provides us now with the information necessary to check the possibility hinted at the beginning of this subsection. Interestingly enough, the minimum time configuration of the generalized Levinson algorithm is the better one for non-symmetric  $L_m$ . Thus, while the storage requirement is the same as that of the corresponding Trench algorithm, it requires  $m^2$  fewer "operations."

<sup>16</sup>The  $g_i$  matrices are needed only to compute the  $\hat{e}_i$ 's and so do not have to be preserved.

**Table 8. Solution of  $k$  sets of equations with common  $L_m$  ( $m$  large)**

Algorithm	Non-symmetric $L_m$		Symmetric $L_m$	
	"Operations"	Storage	"Operations"	Storage
Generalized Levinson (minimum time)	$(2 + k)m^2$	$\frac{1}{2}m^2$	$(1 + k)m^2$	$\frac{1}{2}m^2$
Trench (minimum time)	$(3 + k)m^2$	$\frac{1}{2}m^2$	$(1.5 + k)m^2$	$\frac{1}{4}m^2$
Generalized Levinson (minimum storage)	$3km^2$	$2m$	$2km^2$	$m$
Trench (minimum storage)	$2(1 + k)m^2$	$2m$	$(1 + 1.5k)m^2$	$m$

The symmetric case (still the minimum time configuration) provides an equivocal answer since the modified Levinson algorithm requires  $m^2/2$  fewer "operations" but  $m^2/4$  more cells than the corresponding Trench algorithm.

When storage is at a premium, forcing the adoption of the minimum storage configuration, we have the following situation. For  $k = 2$ , speed and storage requirements are identical for the Trench and generalized Levinson algorithms. However, as  $k$  increases above 2, the Trench algorithm progressively becomes the more efficient one, requiring  $(k - 2)m^2$  fewer "operations" in the non-symmetric case and  $[(k - 1)/2]m^2$  fewer "operations" in the symmetric case.

## References

1. Zohar, S., "Toeplitz Matrix Inversion: The Algorithm of W. F. Trench," *J. Assn. Comp. Mach.*, Vol. 16, No. 4, pp. 592-601, Oct. 1969.
2. Levinson, N., "The Wiener RMS Error Criterion in Filter Design and Prediction," *J. Math. Phys.*, Vol. 25, No. 4, pp. 261-278, Jan. 1947.

## L. Multiple Access Communications Research: Correlation With PN Sequences,

L. D. Baumert and G. A. Morris, Jr.

### 1. Introduction

A computation which frequently arises in digital communications systems is that of finding the correlations between a given data sequence and some, or all, of the cyclic shifts of a PN sequence (i.e., maximal length linear recurring sequence). The purpose of this article is to point out, when more than just a few (see below for a precise

bound) such correlations are to be performed, that there exists a method which is more efficient than direct computation. (The properties of PN sequences used below are established in Ref. 1.)

## 2. Method

While PN sequences are usually thought of as sequences of zeros and ones which are operated on modulo 2, it is more in keeping with the actual digital systems computation to consider them as sequences of ones and minus ones. Thus, the PN sequence 1, 1, 0, 1, 0, 0, 1 is replaced by the sequence -1, -1, 1, -1, 1, 1, -1.

If all cyclic shifts of the PN sequence are to be correlated against the data vector  $x$ , this amounts to the matrix computation

$$Cx = y \quad (1)$$

where  $y$  is the vector of correlation coefficients and  $C$  is the circulant matrix given by

$$C = \begin{bmatrix} a_p & a_{p-1} & \cdot & \cdot & \cdot & a_2 & a_1 \\ a_1 & a_p & \cdot & \cdot & \cdot & a_3 & a_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{p-1} & a_{p-2} & \cdot & \cdot & \cdot & a_1 & a_p \end{bmatrix}$$

Here  $a_p, a_{p-1}, \dots, a_2, a_1$  is the 1, -1 representation of the PN sequence whose period  $p$  is  $2^n - 1$ .

Let  $B$  be the matrix formed by adjoining a row and column of +1's to  $C$ ; that is, let

$$B = \begin{bmatrix} 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 1 & & & & & \\ \cdot & & C & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ 1 & & & & & \end{bmatrix}$$

then

$$B \begin{bmatrix} 0 \\ x \end{bmatrix} = \begin{bmatrix} c(0) \\ y \end{bmatrix} \quad (2)$$

Furthermore, it is known (see proof below) that the matrix  $B$  is equivalent to the matrix  $H$  of the first-order Reed-Müller ( $2^n, n$ ) code. That is, it is known that there exist permutation matrices  $P, Q$  such that

$$H = PBQ$$

So  $y$  can be computed by means of the formula

$$P^{-1} H Q^{-1} \begin{bmatrix} 0 \\ x \end{bmatrix} = \begin{bmatrix} c(0) \\ y \end{bmatrix} \quad (3)$$

that is,  $y$  can be computed in the three steps

$$w = Q^{-1} \begin{bmatrix} 0 \\ x \end{bmatrix} \quad (i)$$

$$z = Hw \quad (ii)$$

$$\begin{bmatrix} c(0) \\ y \end{bmatrix} = P^{-1} z \quad (iii)$$

Since the Computations (i) and (iii) are essentially trivial (given  $P$  and  $Q$ ), the determination of  $y$  amounts to very little more than

$$z = Hw \quad (4)$$

where  $H$  is the matrix of the first-order Reed-Müller ( $2^n, n$ ) code, in 1, -1 notation. SPS 37-53, Vol. III, pp. 185-187, discusses an algorithm for doing the Computation (4) by means of  $n2^n$  operations, whereas the direct computation of  $r$  such correlation coefficients involves  $r(2^n - 1)$  operations. So, whenever

$$r > n \left( \frac{2^n}{2^n - 1} \right) \approx n$$

the algorithm of SPS 37-53, Vol. III, pp. 185-187, should be considered as a method of computing the  $r$  components of  $y$  actually desired.

For  $n = 3, \dots, 9$  and for a particular PN sequence in each case, permutations  $P^{-1}, Q^{-1}$  have been determined.

### 3. Example

Consider the PN sequence  $-1, -1, 1, -1, 1, 1, -1$  and the data sequence  $x_7, x_6, \dots, x_1$ . Here  $Cx$  is

$$\begin{bmatrix} -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} x_7 \\ x_6 \\ x_5 \\ x_4 \\ x_3 \\ x_2 \\ x_1 \end{bmatrix} =$$

$$\begin{bmatrix} -x_7 & -x_6 & +x_5 & -x_4 & +x_3 & +x_2 & -x_1 \\ -x_7 & -x_6 & -x_5 & +x_4 & -x_3 & +x_2 & +x_1 \\ x_7 & -x_6 & -x_5 & -x_4 & +x_3 & -x_2 & +x_1 \\ x_7 & +x_6 & -x_5 & -x_4 & -x_3 & +x_2 & -x_1 \\ -x_7 & +x_6 & +x_5 & -x_4 & -x_3 & -x_2 & +x_1 \\ x_7 & -x_6 & +x_5 & +x_4 & -x_3 & -x_2 & -x_1 \\ -x_7 & +x_6 & -x_5 & +x_4 & +x_3 & -x_2 & -x_1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{bmatrix}$$

whereas  $Hw$  is

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ x_1 \\ x_3 \\ x_7 \\ x_2 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} =$$

$$\begin{bmatrix} x_1 & +x_3 & +x_7 & +x_2 & +x_4 & +x_5 & +x_6 \\ -x_1 & +x_3 & -x_7 & +x_2 & -x_4 & +x_5 & -x_6 \\ x_1 & -x_3 & -x_7 & +x_2 & +x_4 & -x_5 & -x_6 \\ -x_1 & -x_3 & +x_7 & +x_2 & -x_4 & -x_5 & +x_6 \\ x_1 & +x_3 & +x_7 & -x_2 & -x_4 & -x_5 & -x_6 \\ -x_1 & +x_3 & -x_7 & -x_2 & +x_4 & -x_5 & +x_6 \\ x_1 & -x_3 & -x_7 & -x_2 & -x_4 & +x_5 & +x_6 \\ -x_1 & -x_3 & +x_7 & -x_2 & +x_4 & +x_5 & -x_6 \end{bmatrix} \begin{bmatrix} c(0) \\ y_1 \\ y_2 \\ y_4 \\ y_3 \\ y_7 \\ y_5 \\ y_6 \end{bmatrix}$$

Thus, the permutation  $Q^{-1}$  applied to  $x$  sends  $x_1 \rightarrow x_7$ ,  $x_2 \rightarrow x_4$ ,  $x_3 \rightarrow x_6$ ,  $x_4 \rightarrow x_3$ ,  $x_5 \rightarrow x_2$ ,  $x_6 \rightarrow x_1$ ,  $x_7 \rightarrow x_5$ , and so

is completely described by the sequence 7, 4, 6, 3, 2, 1, 5. The permutation  $P^{-1}$  applied to  $z$  sends rows  $0 \rightarrow 0$ ,  $1 \rightarrow 1$ ,  $2 \rightarrow 2$ ,  $3 \rightarrow 4$ ,  $4 \rightarrow 3$ ,  $5 \rightarrow 7$ ,  $6 \rightarrow 5$ ,  $7 \rightarrow 6$ , and thus may be described by the sequence 0, 1, 2, 4, 3, 7, 5, 6.

### 4. Proof that $H = PBQ$

While it is implicit in the results of Ref. 1 that  $H = PBQ$ , this fact is not explicitly demonstrated there. Since this fact is a crucial step for the computation discussed above, a proof is given here. But first some preliminary information must be recalled.

The first-order Reed-Müller  $(2^n, n)$  code is usually considered to be the vector space over  $GF(2)$  generated by the basis vectors

$$\begin{aligned} g_1 &= 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ \dots \quad 0 \ 1 \ 0 \ 1 \\ g_2 &= 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ \dots \quad 0 \ 0 \ 1 \ 1 \\ &\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ g_n &= 0 \ 0 \ 0 \ \dots \quad 0 \ 1 \ 1 \ 1 \ \dots \ 1 \ 1 \ 1 \end{aligned} \quad (5)$$

(This set of basis vectors is conveniently described by noting that its columns consist of the binary expansions of  $0, 1, 2, \dots, 2^n - 1$ .) Thus, every word of this code can be represented as

$$w = \sum_{i=1}^n c_i g_i \quad (6)$$

where the addition is modulo 2 and  $c_i = 0$  or 1. [The fact that this process produces  $2^n$  distinct words  $w$  follows from the observation that the  $(2^i + 1)$ -st component of  $w$  is 1 if and only if  $c_i = 1$ .]

From the point of view of the algorithm of SPS 37-53, Vol. III, pp. 185-187, another representation of the Reed-Müller  $(2^n, n)$  code is more important, however. Let  $E = (e_{ij})$ ,  $F = (F_{ij})$  be matrices of order  $s, t$ , respectively. Then the Kronecker product  $E \otimes F$  of  $E$  and  $F$  is the matrix of order  $st$  given by

$$E \otimes F = \begin{bmatrix} e_{11} F & e_{12} F & \dots & e_{1s} F \\ e_{21} F & e_{22} F & \dots & e_{2s} F \\ \vdots & \vdots & \ddots & \vdots \\ e_{s1} F & e_{s2} F & \dots & e_{ss} F \end{bmatrix}$$

Let

$$\bigotimes_{i=1}^3 E, \bigotimes_{i=1}^4 E, \dots$$

denote

$$(E \otimes E) \otimes E, [(E \otimes E) \otimes E] \otimes E, \dots$$

respectively. Then, with

$$J = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (7)$$

For this is obviously true for  $J$  and may be established for  $H$  by induction on  $n$ . Further, note that Eq. (8) shows that  $H$  contains the following rows ( $r_i$  denotes row  $2^{i-1} + 1$  of  $H$ ):

$$\begin{array}{ll} r_1 = 1, -1, & 1, -1, 1, -1, \dots & 1, -1, & 1, -1 \\ r_2 = 1, & 1, -1, -1, 1, & 1, \dots & 1, & 1, -1, -1 \\ \vdots & & & \\ r_n = 1, & 1, & 1, \dots & 1, -1, -1, \dots & -1, -1, -1, -1 \end{array} \quad (10)$$

which are the 1, -1 representations of  $g_1, \dots, g_n$ , respectively.

Thus, the correspondence between  $H$  and the 0, 1 representation of the Reed-Müller code is given by

$$\sum_{i=1}^n c_i g_i \Leftrightarrow \prod_{i=1}^n r_i^{c_i} \quad (11)$$

where the summation is modulo 2 and the product is that defined by Eq. (9).

Proof that  $H = PBQ$ : First, map the rows  $r_1, r_2, \dots, r_n$  (Eq. 10) of  $H$  into rows 2, 3,  $\dots, n+1$ , respectively, of  $B$ . Thus,  $n$  rows of  $P$  have been determined. Furthermore,  $Q$  is completely determined by this. This is because every 1, -1  $n$ -tuple occurs precisely once as a column in Eq. (10) and, similarly, since rows 2, 3,  $\dots, n+1$  correspond to  $n$  consecutive cyclic shifts of a PN sequence, every 1, -1  $n$ -tuple occurs precisely once as a column among these rows.

By the cycle-and-add property of PN sequences, every product, in the sense of Eq. (9), of rows 2, 3,  $\dots, n+1$

it is known that

$$H = \bigotimes_{i=1}^n J \quad (8)$$

where  $H$  is the 1, -1 representation of the Reed-Müller code (i.e., as in Eq. 4 above). To see that this last statement is true, note that every product of rows of  $H$  is a row of  $H$  under the following multiplication rule:

$$(h_{i_1}, h_{i_2}, \dots, h_{i_m}) \cdot (h_{j_1}, h_{j_2}, \dots, h_{j_m}) = (h_{i_1}h_{j_1}, h_{i_2}h_{j_2}, \dots, h_{i_m}h_{j_m}) \quad (9)$$

of  $B$  is a row of  $B$ . Furthermore, there are  $2^n$  distinct such products, because there are  $2^n$  distinct products of  $r_1, \dots, r_n$ . Thus, every row of  $B$  can be written as a product of rows 2, 3,  $\dots, n+1$ . Thus,  $P$  is completely defined by making correspond to any product of rows 2, 3,  $\dots, n+1$  of  $B$  the associated product of  $r_i$ 's.

## 5. $P$ and $Q$ are not Unique

That  $P$  and  $Q$  are not unique is immediate from the proof given above. For example, rows 3,  $\dots, n+2$  of  $B$  could have been used equally as well as rows 2,  $\dots, n+1$ . Suppose

$$H = P_1 B Q_1 = P_2 B Q_2 \quad (12)$$

then

$$B = P_1^{-1} P_2 B Q_2 Q_1^{-1} \quad (13)$$

Thus,  $B$  is transformed into itself by the row and column permutations  $R = P_1^{-1} P_2$  and  $S = Q_2 Q_1^{-1}$ . This pair of permutations  $R, S$  thus determines an *automorphism* of the matrix  $B$ . On the other hand, any such permutation

automorphism  $R, S$  of  $B$  yields permutation matrices  $P = P_1 R, Q = S Q_1$  which take  $B$  into  $H$ . Thus, the number of distinct permutation pairs  $P, Q$  taking  $B$  into  $H$  is exactly the same as the number of permutation automorphisms  $R, S$  of  $B$ . The number of such automorphisms is known (Ref. 2) to be at least

$$2^{n(n-1)/2} \prod_{i=2}^n (2^i - 1)$$

Thus, for example, when  $n = 5$  there are at least 9,999,360 different transformations  $P, Q$  taking  $B$  into  $H$ . Which one to use depends on the simplicity of the necessary logic.

### References

1. Golomb, S. W., *Shift Register Sequences*. Holden-Day, Inc., San Francisco, Calif., 1967.
2. Dembowski, P., *Finite Geometries*. Springer-Verlag, Berlin, Germany, 1968.

## M. Permutation Transformation of Audio,

R. M. Goldstein and E. C. Posner

### 1. Description

This work involves the demonstration of a novel concept for disguising speech to make it incomprehensible to an unauthorized listener, even with the same receiver as an authorized listener. *Subsection 2* describes the system concept. Previous methods for disguising speech have suffered from at least one of two drawbacks. First, the speech bandwidth was vastly expanded beyond the bandwidth available. The device proposed here is expected to require little bandwidth expansion because of its analog nature. Second, those systems which did not expand bandwidth did not disguise speech, even to a relatively unskilled unauthorized listener. The proposed system, because of the novel and nonlinear nature of the disguise, is expected to prevail in field use even if a relatively skilled listener has the same equipment as the authorized listener, but lacks the key. It is not, however, expected to prevail against a listener with vast resources, including technology, time, and money.

### 2. System Description

The idea is to perform a "running permutation" of an input speech signal, and then transmit the permuted values as an analog signal. Although the permuting is

done digitally, the sequence of permuted samples is converted to an analog signal before transmission. The subsequent discussion assumes that the analog samples themselves are permuted.

If speech is low-pass filtered to 2.5 kHz, then it can be sampled at 5 kHz with no loss of fidelity. It is these samples that are permuted. After permutation, the samples are low-pass filtered at 2.5 kHz, without loss of fidelity, and then transmitted. This results in a transmitted signal of the same bandwidth as the input signal.

Figure 8 shows a block diagram of the permutation system transmitter. Here  $n$ , the number of stages in the storage register, is 3; for real security,  $n$  would presumably be about 30, although this value needs to be determined. It depends on how low a frequency one allows in the speech input. Having chosen  $n$ , we choose a sequence of the integers from 1 to  $n$ , the "key," known to transmitter and receiver. The only restriction on the sequence is that every symbol from 1 to  $n$  occurs in every string of  $2n - 1$  symbols of the key. In the example, the key sequence is

1 3 2 2 1 3 2 1 3 1 2 3 . . .

Let the speech samples in their original order be  $x_1, x_2, x_3, \dots$ . When  $x_1$  arrives, the key is "1", so  $x_1$  is put into the first (leftmost) position in the register. Nothing is yet transmitted. The sample  $x_2$  is put into the third position, because the second key symbol is 3. Similarly,  $x_3$  is put into the second position because the third key symbol is 2. Nothing has yet been transmitted. However, the fourth key symbol is 2, so  $x_4$  is put into the second position, forcing out the  $x_3$  in the second position, which is now transmitted. The fifth key symbol is 1, so  $x_5$  is put into the first position, transmitting  $x_1$ . Continuing, we put  $x_6$  in the third position, transmitting  $x_2$ , etc., according to Table 9.

At the receiver, the process can be reversed since the key sequence is known. In general, if  $y_j$  represents the  $j$ th received sample, find the key value of the received symbol in Table 9. The corresponding index of the transmitted sample is then the key index of the previous occurrence of that key value. Thus,  $y_1$  yields key value 2. The preceding 2 above it has key index 3, so  $x_3$  was first transmitted. Similarly,  $y_2$  gives key value 1, and the previous 1 has key index 1, so  $x_1$  is the second sample transmitted.

The actual mechanization might be quite different, but the mathematical operations will be as described above.

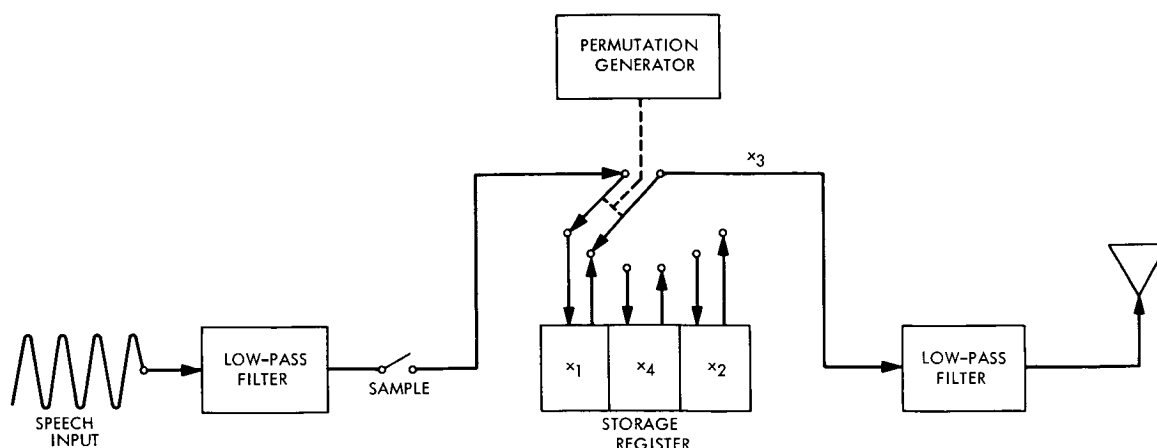


Fig. 8. Permutation system transmitter

Table 9. Key sequence

Key index	Key value	Transmitted	Received value
1	1		
2	3		
3	2		
4	2	$x_3$	$y_1$
5	1	$x_1$	$y_2$
6	3	$x_2$	$y_3$
7	2	$x_4$	$y_4$
8	1	$x_5$	$y_5$
9	3	$x_6$	$y_6$
10	1	$x_8$	$y_7$
11	2	$x_7$	$y_8$
12	3	$x_9$	$y_9$
.	.	.	.
.	.	.	.
.	.	.	.

### 3. Plan

It is planned that this effort will prove the feasibility of this system. The following problems need solution:

- (1) It must be demonstrated that speech can be masked by the permutation device with register lengths  $n$  short enough to be built.
- (2) It must be demonstrated that the authorized recovery produces speech of good quality with little bandwidth expansion.
- (3) Devices for generating random permutations must be found, as well as the theory behind the devices; the next article (Section N) gives one such device, with the theory to go with it.

- (4) The structure of the permutations must be studied to obtain classes of permutations easy to generate but hard to distinguish from each other by an unauthorized listener.
- (5) The synchronization of transmitter key and receiver key must be devised and demonstrated.

Of these five problems, (1) and (2) will be done using computer simulation. Item (5), synchronization, will require some hardware demonstration. Items (3) and (4) involve mathematical ideas from group theory.

## N. A Class of Sequence Permutations, L. R. Welch<sup>17</sup>

### 1. Introduction

The previous article calls for a sequence-permutation device; this article presents one. A device which produces sequence permutations with reasonable properties of randomness must have two features: (1) internal storage to effect the time delay of data terms, and (2) a method of using a random source to determine the time delays of the various data terms. When there is a serial flow of data, the simplest forms of storage are shift registers and delay lines. The storage in the device described below will be referred to as delay lines although it can be mechanized with shift registers. The simplest sources of randomness produce binary sequences, and it will be assumed that such sources are available as input to the permutor.

Another desirable feature at the permutor is that another simple device can use the same random binary

<sup>17</sup>Consultant, Electrical Engineering Department, University of Southern California, Los Angeles, California.

sequence to effect the inverse permutation. The device to be described will have that feature.

## 2. Simple Example

Let  $\{x(t) : t = 1, 2, \dots\}$  be a data sequence and  $\{c(t) : t = 1, 2, \dots\}$  be a control sequence of zeroes and ones. Define the contents of a storage element  $S$  and an output  $Y$  as follows:

$$\left. \begin{aligned} Y(t) &= c(t)X(t) + \bar{c}(t)S(t-1) \\ S(t) &= c(t)S(t-1) + \bar{c}(t)X(t) \end{aligned} \right\} \quad (1)$$

The sequence  $\{Y(t)\}$  is then the permuted sequence.

The output may be directly expressed as a function of the input. If  $t$  is repeatedly decreased in the second equation and the result substituted into the first, the result is

$$\begin{aligned} Y(t) &= c(t)X(t) \\ &+ \bar{c}(t) \sum_{k=1}^t \left[ \prod_{d=1}^k c(t-d) \right] \bar{c}(t-k)X(t-k) \\ &+ \bar{c}(t) \left[ \prod_{d=1}^t c(t-d) \right] S(0) \end{aligned} \quad (2)$$

Equation (2) expresses  $Y(t)$  as a linear combination of terms of the  $X$  sequence and a constant. Since no more than one of the coefficients can be non-zero,  $Y(t)$  is a selection of one of the terms in the set  $\{X(d) : d = 1, \dots, t\}$ , or  $S(0)$ . With no restrictions on the control sequence, arbitrarily long delays can be produced by the device. Since this unduly complicates the design of the inverse device, some conditions on the control sequence are necessary. The simplest condition is  $c(t)c(t-1) \equiv 0$ . Then Eq. (2) reduces to

$$\begin{aligned} Y(t) &= c(t)X(t) + \bar{c}(t)\bar{c}(t-1)X(t-1) \\ &+ c(t-1)X(t-2) \end{aligned} \quad (3)$$

Since  $c(t)$  and  $c(t-1)$  cannot both equal 1, it is seen that  $Y(t)$  is selected from among  $X(t)$ ,  $X(t-1)$ ,  $X(t-2)$ . Figure 9a illustrates the device. The control sequence generation is discussed later.

Next define  $Z, D$  by

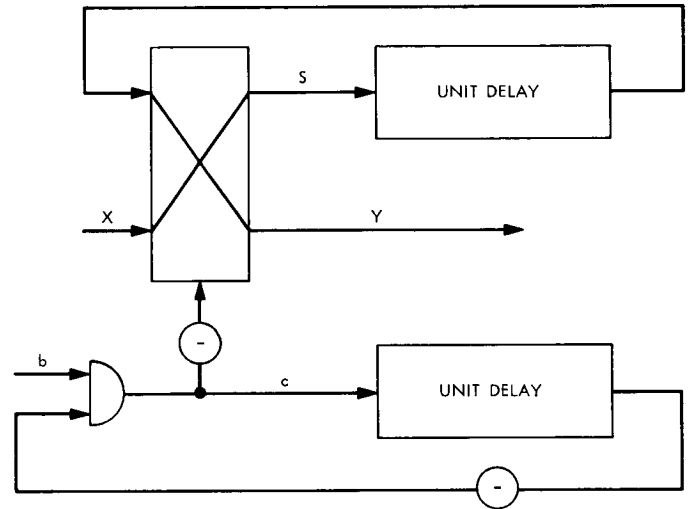
$$\left. \begin{aligned} Z(t) &= c(t-1)Y(t) + \bar{c}(t-1)D(t-1) \\ D(t) &= \bar{c}(t-1)Y(t) + c(t-1)D(t-1) \end{aligned} \right\} \quad (4)$$

By analogy with Eq. (3), it follows that

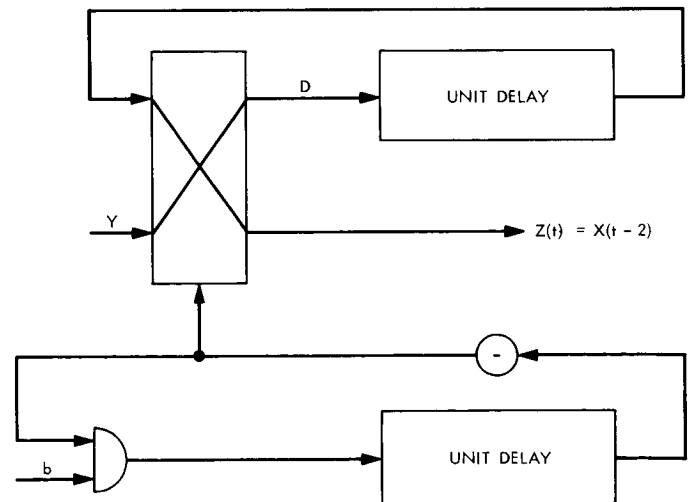
$$\begin{aligned} Z(t) &= c(t-1)Y(t) + \bar{c}(t-1)\bar{c}(t-2)Y(t-1) \\ &+ c(t-2)Y(t-2) \end{aligned} \quad (5)$$

If Eq. (3) is used to eliminate  $Y(t)$ ,  $Y(t-1)$ , and  $Y(t-2)$ , it is found that  $Z(t) = X(t-2)$ . Therefore, every data term is present in the  $Y$ -sequence, and Eq. (3) represents a permutation. Furthermore, the inverse permutation (followed by a delay of two units of time) can be effected by the same device, using the delayed control sequence. Figure 9b illustrates the inversion.

(a) PERMUTOR



(b) DE-PERMUTOR



SWITCH  ACTS AS AN INTERCHANGE BETWEEN UPPER AND LOWER INPUTS WHEN BOTTOM INPUT IS 1

Fig. 9. Single-stage permutor

There are two simple methods of producing the control sequence from a random binary stream  $\{b(t)\}$ . The first is  $c(t) = \bar{b}(t) b(t-1)$ . It is easily seen that  $c(t) c(t-1) \equiv 0$ . Also,

$$\left. \begin{aligned} \text{prob}[Y(t) = X(t)] &= \text{prob}[Y(t) = X(t-2)] = \frac{1}{4} \\ \text{prob}[Y(t) = X(t-1)] &= \frac{1}{2} \end{aligned} \right\} \quad (6)$$

The second method is to define  $c(t) = b(t) \bar{c}(t-1)$ . Again it is easily seen that  $c(t) c(t-1) \equiv 0$ . The sequence  $\{c(t)\}$  is a binary markov chain, and, with the stationary initial probabilities,

$$\text{prob}[Y(t) = X(t-d)] = \frac{1}{3} \quad \text{for } d = 0, 1, 2$$

This method is used in the simple implementation of Figs. 9 and 10. Figure 10 will be explained in the next subsection.

### 3. General Model

The simple model does not displace data samples more than two time units. This can be modified by replacing the unit delay by some other constant. Furthermore, a more sophisticated device results if several simple devices are arranged in series. This is illustrated in Fig. 10 and in the following system of equations. Let  $N_1, \dots, N_k$  be positive integers and  $\{b_1(t)\}, \dots, \{b_k(t)\}$  be mutually

independent Bernoulli sequences. Define the system

$$\left. \begin{aligned} Y_0(t) &= X(t) \\ c_i(t) &= b_i(t) \bar{c}_i(t - N_i) \\ Y_i(t) &= c_i(t) Y_{i-1}(t) + \bar{c}_i(t) S_i(t - N_i) \\ S_i(t) &= c_i(t) S_i(t - N_i) + \bar{c}_i(t) Y_{i-1}(t) \\ Y(t) &= Y_k(t) \end{aligned} \right\} \quad (7)$$

for  $i = 1, 2, \dots, k$ . Then the terms of  $Y$  are the terms of  $X$  in permuted order.

The inverse permutation has an analogous description, complicated by the fact that the  $j$ th stage and its inversion produces a net delay of  $2N_j$  and the stages must be applied in the reverse order. The system is (not illustrated):

$$\left. \begin{aligned} z_k(t) &= Y(t) \\ M_j &= 2 \sum_{i=1}^{j-1} N_{k+1-i} \\ d_{k-j}(t) &= b_{k+1-j}(t - M_j) \bar{d}_{k-j}(t - N_{k+1-j}) \\ z_{k-j}(t) &= d_{k-j}(t - N_{k+1-j}) z_{k-j+1}(t) \\ &\quad + \bar{d}_{k-j}(t - N_{k+1-j}) D_{k-j}(t - N_{k+1-j}) \\ D_{k-j}(t) &= d_{k-j}(t - N_{k+1-j}) D_{k-j}(t - N_{k+1-j}) \\ &\quad + \bar{d}_{k-j}(t - N_{k+1-j}) z_{k-j+1}(t) \\ X(t - M_{k+1}) &= z_0(t) \end{aligned} \right\} \quad (8)$$

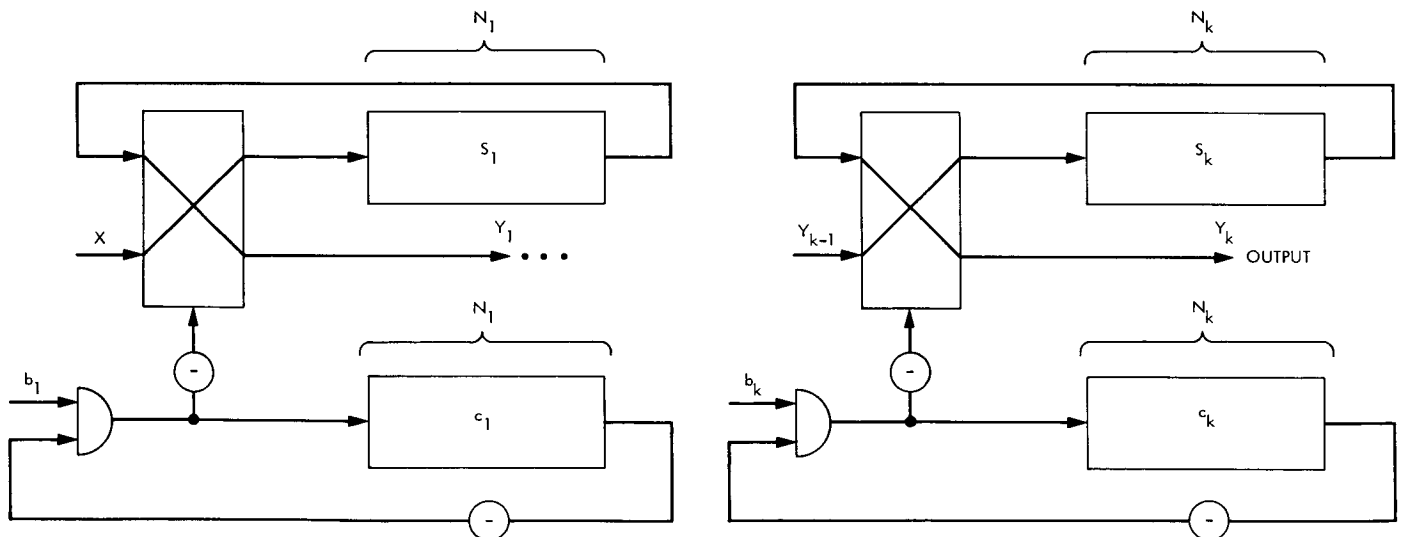


Fig. 10. Multi-stage permutor

#### 4. Statistics

From the system of Eqs. (8), it is possible to represent  $X$  as

$$X(t) = Y[t + T(t)]$$

where  $T$  is a function of the control sequences and its value is the amount of time the term  $X(t)$  is delayed.  $T(t)$ , in its dependence on the random sequences, is a stochastic sequence. Its first-order distribution can be computed as

$$\text{prob}(T_j = 0) = \text{prob}(T_j = N_j) = \text{prob}(T_j = 2N_j) = \frac{1}{3}$$

when the control sequences are generated by Eq. (7), and the characteristic function of  $T_j$  is

$$\left. \begin{aligned} \phi_j(\lambda) &= \frac{1}{3} [\exp c\lambda N_j] (1 + 2 \cos \lambda N_j) = \frac{1}{3} \frac{\sin \frac{3}{2} \lambda N_j}{\sin \frac{1}{2} \lambda N_j} \exp(i N_j) \\ \phi_T(\lambda) &= \left(\frac{1}{3}\right)^k \frac{\prod_{j=1}^k \sin \frac{3}{2} \lambda N_j}{\prod_{j=1}^k \sin \frac{1}{2} \lambda N_j} \exp\left(i \sum_{j=1}^k N_j\right) \end{aligned} \right\} \quad (9)$$

In the case where the  $N_j = 3^{j-1}$ , there is a cancellation of terms in Eq. (9) and the resulting characteristic function is that of a uniform distribution on the integers from 0 to  $3^k - 1$ . This is a very desirable property.

The joint distribution of  $[T(t), T(t + \tau)]$  is not so simply arrived at. That they are not independent can be seen from

$$t + T(t) \neq \tau + t + T(t + \tau) \quad \text{for } \tau \neq 0$$

However, the degree of dependence is much more than this indicates. In the case where  $N_j = 3^{j-1}$ , the control bits can be recovered from the sequence  $\{T(1), T(2), \dots\}$  with bounded time delay. Therefore, the entropy rate of  $T(t)$  is equal to the joint entropy rate of the control sequences. The entropy rate of a control sequence is  $2/3 \ln 2$ , so that  $R_T = 2/3 k \ln 2$ . This contrasts with the monographic entropy  $R_1 = k \ln 3$ .  $(R_1 - R_T)$  divided by  $R_1$  is a measure of redundancy and equals  $1 - \log_{27} 4$ . This is somewhat greater than one-half. The best sequence would have redundancy zero.

follows. First,

$$T(t) = T_1(t) + T_2(t) + \dots + T_k(t)$$

where  $T_j(t)$  is the delay introduced by the  $j$ th stage. Now, while it is true that the first delay affects which control terms are used in the second stage and thus affects the second delay, the probability distribution at the second stage is unaffected. Extending this observation, we find that  $T_1, \dots, T_k$  are independent random variables. Now,

## O. Spectrum Analysis of Scrambling Permutation,

U. Timor

### 1. Introduction

The previous article gives a method for disguising speech to make it incomprehensible to an unauthorized listener. If a speech is low-pass filtered to 2.5 kHz, it can (ideally) be sampled at 5 kHz and reconstructed from the samples with no loss of fidelity. If the samples are permuted, and if the permuted sequence of samples is converted to an analog signal before transmission, the output signal has the same bandwidth (2.5 kHz) as the original speech, and is disguised. At the receiver, the process is reversed to yield the original signal.

The key is a sequence of integers from 1 to  $n$ . The effectiveness of the scrambling depends on the structure of the key sequence, and the effectiveness of the best sequence of length  $n$  increases with  $n$ . On the other hand, the complexity of the system increases (linearly) with  $n$ , so we want to find the minimum  $n$  which yield an accepted

scrambling. One good property is to demand that the spectrum of the scrambled speech be reasonably white. This article studies what happens to a sine wave through the system. If a sine wave comes out reasonably white, it is to be expected that speech, which is composed of sine waves, will come out white, too.

The purpose of this work is to find the amount of scrambling as a function of  $n$ , the "constraint length." The analysis will be conducted both in the frequency domain and the time domain. First, we will compare the spectrum of a known input signal  $[x(t) = \cos w_0 t]$  with the spectrum of its permuted waveform. As it turns out, the effect of the scrambling is to spread the energy almost homogeneously over the band of transmission, leaving only a fraction  $\epsilon(n)$  of the energy around  $w_0$ , where  $\epsilon(n)$  decreases to zero with  $n$ . A second measure of scrambling will be the time average of the square of the difference between the input and output signal.

The key was taken to be a Bernoulli sequence on  $n$  symbols, i.e., in independent random sequence from an  $n$ -symbol alphabet. As will be indicated in Subsection 6, such a sequence requires an infinite storage of the receiver, and thus cannot be a key. Therefore, the analysis can be considered only as an approximation to a practical system.

## 2. Description of the System

Let  $x(t)$  be the input signal which is assumed to have a bandwidth of  $W$  ( $W = 2.5$  kHz). By the sampling theorem (Ref. 1),  $x(t)$  can be represented exactly by its samples at  $k/2W$ ,  $k = 0, \pm 1, \pm 2, \dots$

$$x(t) = \sum_{k=-\infty}^{\infty} x_k v\left(t - \frac{k}{2W}\right)$$

where

$$x_k = x\left(\frac{k}{2W}\right), \quad v(t) = \frac{\sin(2\pi Wt)}{2\pi Wt}$$

Let the key sequence be  $a_1, a_2, a_3, \dots$ , where each  $a_i$  is an integer between 1 and  $n$ . The transmitter has  $n$  storage cells (numbered from 1 to  $n$ ) and operates as follows: At time  $t = 1/2W$ ,  $x_1$  is stored at cell  $a_1$ ; at  $t = 2/2W$ ,  $x_2$  is stored at cell  $a_2$ , and so on. Since there are  $n$  storage cells, after at most  $n$  samples we will have some  $x_k$  with  $a_k = a_l$ ,  $l < k$ . That is, cell  $a_k$  already contains the sample  $x_l$ . In this case, while  $x_k$  is stored,  $x_l$  will be the output  $y_k$ .

Let  $d_k$  denote the difference of indices between  $a_k$  and  $a_l$ . That is, if for  $k > l$ , we have

$$a_k = a_l$$

$$a_i \neq a_k, \quad i = l + 1, \dots, k - 1$$

then

$$d_k = k - l$$

After a short starting period, we will have an output sequence  $y_k, y_{k+1}, y_{k+2}, \dots$ , where  $y_k = x_{k-d_k}$ . The transmitted analog signal will be

$$v(t) = \sum_{k=-\infty}^{\infty} y_k v\left(t - \frac{k}{2W}\right) = \sum_{k=-\infty}^{\infty} x_{k-d_k} v\left(t - \frac{k}{2W}\right)$$

which depends on the input signal and the key sequence (through  $\{d_k\}$ ). At the receiver,  $y(t)$  is again sampled at  $t = k/2W$ , and  $\{y_k\}$  is repermuted according to the key to yield  $\{x_k\}$ , from which  $x(t)$  can be reconstructed with no loss of fidelity.

## 3. Statistics of $d$

Let the key be a sample sequence of a Bernoulli sequence of index  $n$ , i.e., the key is  $a_1, a_2, \dots$ , where

- (1)  $a_i \in [1, 2, \dots, n]$ .
- (2) all  $a_i$  are independent.
- (3)  $\text{prob}(a_i = k) = \frac{1}{n}$ ,  $k = 1, 2, \dots, n$ .

As described in Subsection 2, we can define a corresponding shift sequence  $\{d_k\}$ , where  $d_k$  is a positive integer. We have the following statistical properties of  $\{d_k\}$ :

- (1)  $P_n(m) \triangleq \text{prob}[d_k = m] = \frac{1}{n} \left(\frac{n-1}{n}\right)^{m-1}$ ,  
 $m = 1, 2, \dots$ , independent of  $k$ .
- (2)  $E\{d\} = \sum_{m=1}^{\infty} m P_n(m) = n$ .
- (3)  $\sigma_d^2 = n^2 - n$ .
- (4)  $\text{prob}\{d \geq L\} = \left(\frac{n-1}{n}\right)^{L-1}$ .
- (5) The joint probability  $p_n(d_k, d_l)$  depends on the difference  $k - l$ .

If  $k = l$ ,

$$p_n(d_k, d_k) = p_n(d_k)$$

If  $k > l$ ,

$$p_n(d_k, d_l) = \begin{cases} p_n(d_k) p_n(d_l), & d_k \leq k - l \\ p_n(d_k) p_{n-1}(d_l), & d_k > k - l \text{ and } d_l < d_k - (k - l) \\ 0, & d_k > k - l \text{ and } d_l = d_k - (k - l) \\ \left[ \frac{(n-1)^2}{n(n-2)} \right]^{k-l} p_{n-1}(d_k) p_n(d_l), & d_k > k - l \text{ and } d_l > d_k - (k - l) \end{cases}$$

where

$$p_{n-1}(d) = \frac{1}{n-1} \left( \frac{n-1}{n-2} \right)^{d-1}$$

Similarly for  $k < l$ .

#### 4. Spectra of Input and Output Signals

In spectral analysis of speech (Ref. 2), we distinguish between short-time spectrum and average (or long-term) spectrum. The vocal mechanism is a quasi-stationary source of sound, and its excitation and normal modes change with time. To reflect significant temporal changes, we have to consider only a portion of the signal extended over the time appropriate to the quasi-steady elements of the speech signal.

This can be achieved by analyzing a portion of the signal "seen" through a specified time window. The length of the time window determines the compromise made between temporal and frequency resolution. A duration of about 8 ms corresponds to a single pitch period of a male voice, and this will yield a spectral analysis in which the temporal structure of individual vocal periods is resolved. A duration of 40 ms covers several pitch

periods and is commonly used to provide a frequency resolution adequate to resolve the harmonic components in the voiced portion of speech. A time window of 20 ms is a sort of time-frequency compromise. All these cases fall under the category of short-time analysis. In long-term analysis, we consider the average spectrum distribution over a long period of speech, rather than short-time variation.

**a. Average power spectrum.** Let the input be  $x(t) = \cos(\omega_0 t)$ ,  $\omega_0 < 2\pi W$ . The output signal is

$$y(t) = \sum_{k=-\infty}^{\infty} \cos[b(k - d_k)] v\left(t - \frac{k}{2W}\right)$$

where

$$b \triangleq \frac{\omega_0}{2W}$$

The average power spectrum  $\phi_y(f)$  is the Fourier transform of the long-term correlation function  $\varphi_y(\tau)$ ,

$$\varphi_y(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T y(t) y(t + \tau) dt$$

We approximate  $y(t)$  in  $[-T, T]$  by taking only the samples within this interval, i.e.,  $|k|/2W \leq T$ .

$$\varphi_y(\tau) \approx \lim_{T \rightarrow \infty} \frac{1}{2T} \sum_{k=-[2WT]}^{[2WT]} \sum_{l=-[2WT]}^{[2WT]} \cos[b(k - d_k)] \cos[b(l - d_l)] \int_{-T}^T v\left(t - \frac{k}{2W}\right) v\left(t + \tau - \frac{l}{2W}\right) dt$$

But

$$\int_{-T}^T v\left(t - \frac{k}{2W}\right) v\left(t + \tau - \frac{l}{2W}\right) dt \approx \int_{-\infty}^{\infty} v\left(t - \frac{k}{2W}\right) v\left(t + \tau - \frac{l}{2W}\right) dt = \frac{1}{2W} v\left(\tau + \frac{k-l}{2W}\right)$$

Hence, if we denote  $M = [2WT]$ , then

$$\varphi_y(\tau) \approx \lim_{M \rightarrow \infty} \frac{1}{2M} \sum_{k=-M}^M \sum_{l=-M}^M \cos[b(k - d_k)] \cos[b(l - d_l)] v\left(\tau + \frac{k-l}{2W}\right)$$

Taking the Fourier transform of  $v[\tau + (k-l)/2W]$  and combining terms of  $k, l$  with  $k-l = \pm m$ , we get

$$\phi_y(f) = \lim_{M \rightarrow \infty} \frac{1}{4MW} \left\{ \sum_{k=-M}^M \cos^2[b(k - d_k)] + 2 \sum_{k=-M+1}^M \sum_{l=-M}^{k-1} \cos[b(k - d_k)] \cos[b(l - d_l)] \cos\left[\frac{\pi f}{W}(k-l)\right] \right\}$$

when  $|f| \leq W$ , and  $\phi_y(f) = 0$  for  $|f| > W$ .

Since  $\{d\}$  is a random sequence,  $\phi_y(f)$  is a random process. Its expected value is

$$S_y(f) = \sum_{d_k=1}^{\infty} \sum_{d_l=1}^{\infty} p_n(d_k, d_l) \phi_y(f)$$

Substituting  $p_n(d_k, d_l)$  from Subsection 3 and performing the double summation (Ref. 3), we can write  $S_y(f)$  as follows:

$$S_y(f) = \lim_{T \rightarrow \infty} \frac{1}{8W^2T} \left\{ \sum_{k=-2WT}^{2WT} G(k) + 2 \sum_{k=-2WT+1}^{2WT} \sum_{l=-2WT}^{k-1} \left[ C(k)C(l) - \left(\frac{n-1}{n}\right)^{k-l} D(l) \right] \cos\left[\frac{\pi f}{W}(k-l)\right] \right\}, \quad |f| \leq W$$

where

$$G(k) = \frac{n \sin(b) \sin[b(2k-1)] + \frac{1}{2} \cos(2bk)}{2n(n-1)[1 - \cos(2b)] + 1} + \frac{1}{2}$$

$$C(k) = \frac{2n \sin\left(\frac{b}{2}\right) \sin\left[b\left(k - \frac{1}{2}\right)\right] + \cos(bk)}{\alpha_n}$$

$$\begin{aligned} D(l) = C^2(l) - \left(\frac{n-1}{n}\right) & \frac{\left\{ 2(n-1) \sin\left(\frac{b}{2}\right) \sin\left[b\left(l - \frac{1}{2}\right)\right] + \cos(bl) \right\} \left\{ 2n \sin\left(\frac{b}{2}\right) \sin\left[b\left(l - \frac{3}{2}\right)\right] + \cos[b(l-1)] \right\}}{\alpha_n \alpha_{n-1}} \\ & + \frac{(n-1)(n-2) - (n-2)^2 \cos(b)}{4n\alpha_{n-1}} + \frac{(n-1)(n-2) \sin\left(\frac{b}{2}\right) \left\{ \sin\left[2b\left(l - \frac{3}{4}\right)\right] - n \sin(b) \cos\left[2b\left(l - \frac{5}{n}\right)\right] \right\}}{n\alpha_{n-1}\beta} \\ & + \frac{\left\{ (n-2) \cos\left[2b\left(l - \frac{1}{2}\right)\right] + n \sin(b) \sin[2b(l-1)] \right\}}{2n\alpha_{n-1}\beta} - \frac{\cos(2bl) + n \sin(b) \sin\left[2b\left(l - \frac{1}{2}\right)\right]}{2\alpha_n\beta} \\ & + \frac{n[1 - \cos(b)] - 1}{4\alpha_n} - \frac{n \sin\left(\frac{b}{2}\right) \left\{ \sin\left[2b\left(l - \frac{1}{4}\right)\right] - n \sin(b) \cos\left[2b\left(l - \frac{3}{4}\right)\right] \right\}}{\alpha_n\beta} \end{aligned}$$

and

$$\alpha_n = 2n(n-1)[1 - \cos(b)] + 1$$

$$\beta = n(n-2)[1 - \cos(2b)] + 2$$

Note that all these terms are independent of  $f$ .

**b. Short-time spectrum.** Let the input signal be  $x(t) = \cos(\omega_0 t)$ . We consider only a portion of the signal of duration  $T$ . Thus,

$$F_T(f, t_1) = \int_{t_1-T}^{t_1} \cos(\omega_0 t) \exp(-2\pi f t) dt$$

and the power spectrum is

$$S_x(F, t_1, T) = \frac{1}{T} [F_T(f, t_1)]^2$$

$$= A^2 + B^2 + 2AB \cos \left[ 2\omega_0 \left( t_1 - \frac{T}{2} \right) \right]$$

Let  $M_1 = 2W(t_1 - T)$ ,  $M_2 = 2Wt_1$ ; then for  $|f| \leq W$ ,

$$\frac{1}{T} [F_T(f, t_1)]^2 = \frac{1}{2W(M_1 + M_2)} \left\{ \sum_{k=M_1}^{M_2} \cos^2[b(k - d_k)] + \sum_{k=M_1+1}^{M_2} \sum_{l=M_1}^{k-1} \cos[b(k - d_k)] \cos[b(l - d_l)] \cos \left[ \frac{\pi f}{W} (k - l) \right] \right\}$$

Note that if we put  $M_2 = M$ ,  $M_1 = -M$  and take the limit as  $M \rightarrow \infty$ , we get the long-term spectrum. Therefore, the expected value of the short-time spectrum can be found from the formula of the long-term spectrum, by an appropriate modification of the range of summation.

**c. Numerical results.** Short-time spectra for an input signal  $x(t) = \cos(2\pi f_0 t)$ ,  $f_0 = 300$  Hz and for output signals with different indices of permutation ( $n = 4, 8, 16, 32$ ) were computed for time windows of 8, 20, and 40 ms. Figures 11 and 12 show the results for the case  $T = 20$  ms. To compute the average spectrum, we had to take  $T \rightarrow \infty$  (or an infinite number of samples), which cannot be performed by numerical methods. As an indication, the case of  $T = 0.2$  s was computed.

We can draw the following conclusions:

- (1) The spectrum of the local signal is concentrated in a band  $\pm [f_0 - \Delta f, f_0 + \Delta f]$ , where  $\Delta f \approx 1/T$ . The corresponding output signal has a fraction  $\epsilon$  of its energy in the same band around  $f_0$ , and the remain-

where

$$A = \frac{(T)^{1/2}}{2} \frac{\sin[\pi T(f + f_0)]}{\pi T(f + f_0)}$$

$$B = \frac{(T)^{1/2}}{2} \frac{\sin[\pi T(f - f_0)]}{\pi T(f - f_0)}$$

The corresponding portion of the output signal is

$$y(t) = \sum_{k=-\infty}^{\infty} \cos[b(k - d_k)] v \left( t - \frac{k}{2W} \right),$$

$$t_1 - T \leq t \leq t_1$$

Using the same approximation as for the long-term correlation function, we get

$$F_T(f, t_1) \approx \sum_{2W(t_1-T)}^{2Wt_1} \cos[b(k - d_k)] V(f) \exp \left( -\frac{i\pi f k}{W} \right)$$

ing energy is "white" (i.e., almost equally distributed) in the frequency range  $[-W, W]$ .

- (2) For the cases computed,  $\epsilon$  is almost independent of the time duration of the signal.
- (3)  $\epsilon$  is a decreasing function of the index of permutation  $n$ . For  $f_0 = 300$  Hz, we have

$n$	4	8	16	32
$\epsilon, \%$	37.5	11	2.85	0.7

## 5. The Mean-Square Difference

In comparing the input and output spectra, we have a measure of scrambling in the frequency domain. We can also obtain a measure of scrambling in the time domain. Let the input signal be  $x(t)$ . The output is

$$y(t) = \sum_{k=-\infty}^{\infty} x_{k-d_k} v \left( t - \frac{k}{2W} \right)$$

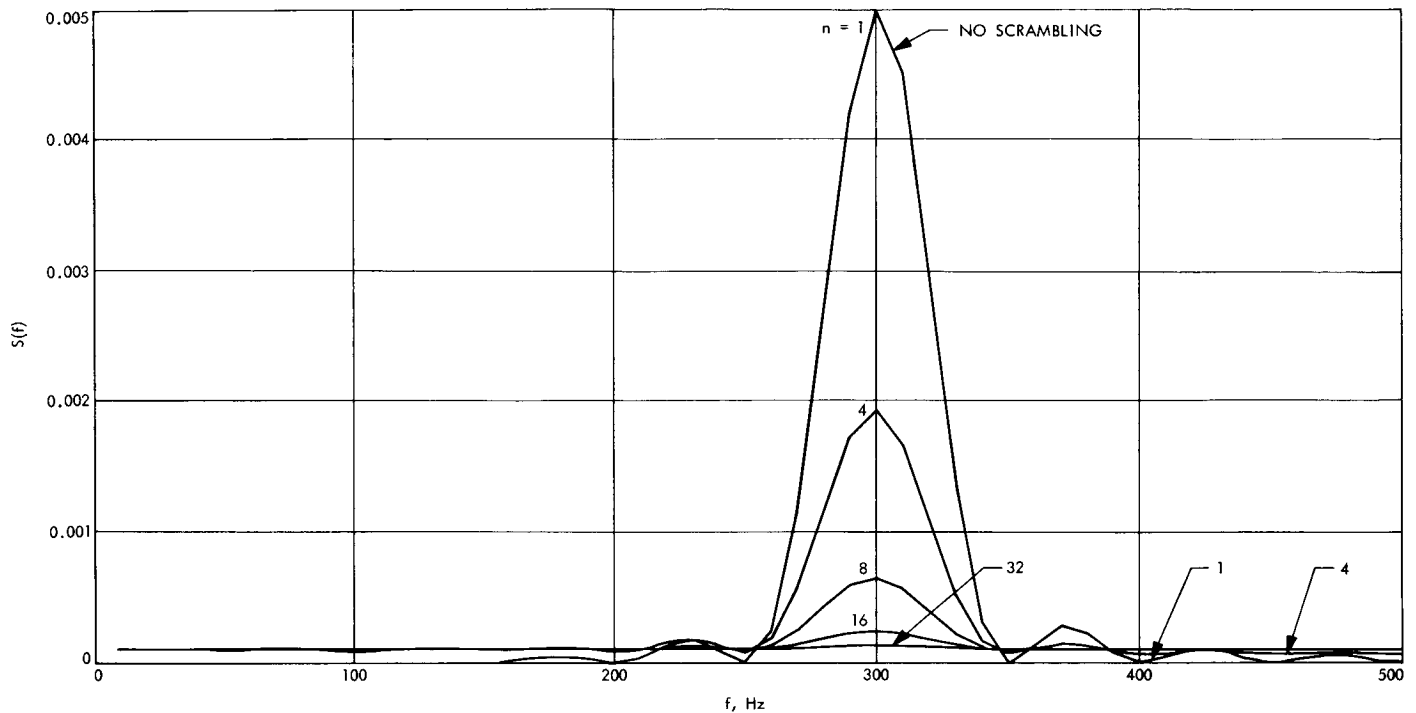


Fig. 11. Spectra for input and output signals [scale chosen to show  $\epsilon(n)$ ]

where each input sample is delayed  $d/2W$  seconds ( $d$  is a random variable). Since  $\bar{d} = n$ , the average delay of an input sample is  $n/2W$  and, therefore, the only meaningful comparison is between  $x(t)$  and  $y[t + (n)/2W]$ . Let

$$z(t) \triangleq y\left(t + \frac{n}{2W}\right) - x(t)$$

$$= \sum_{k=-\infty}^{\infty} (x_{k+n-d} - x_k) v\left(t - \frac{k}{2W}\right)$$

The mean-square difference is the expected value of

$$A = \frac{1}{2T} \int_{-T}^T z^2(t) dt$$

where we can consider a finite  $T$ , or

$$\lim_{T \rightarrow \infty} A$$

(corresponding to short- or long-term difference). Using the same approximation as before, we get

$$A \approx \frac{1}{4WT} \sum_{k=-2WT}^{2WT} (x_{k+n-d} - x_k)^2 = \overline{(x_{k+n-d} - x_k)^2}$$

where the average is taken over  $4WT$  samples. The mean-square difference is

$$MSD = E_d A = \overline{E_d (x_{k+n-d} - x_k)^2}$$

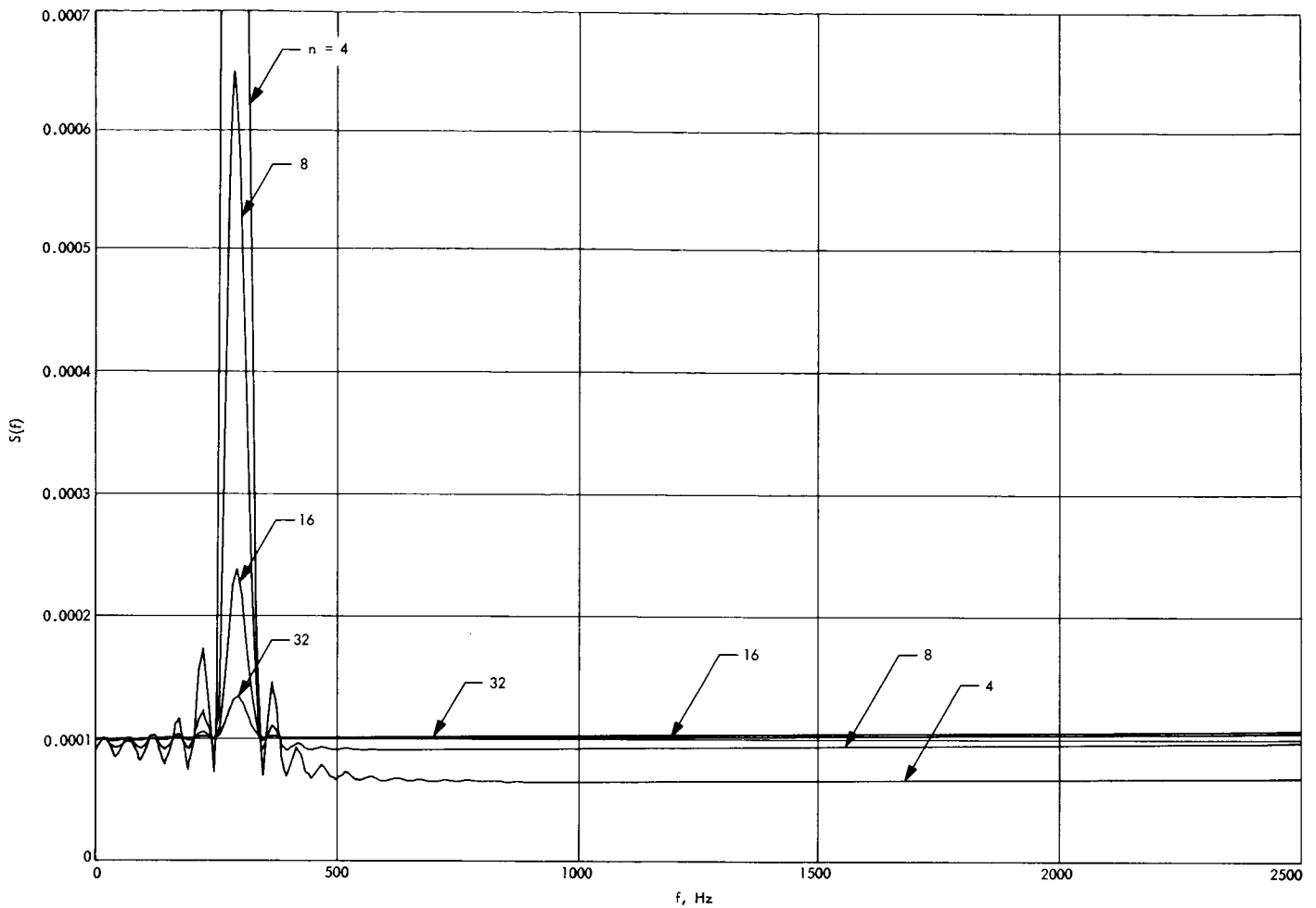


Fig. 12. Spectra for output signals [scale chosen to show the "white" position of  $S(f)$ ]

*Case I.*  $x(t)$  is a (quasi-) stationary ergodic random process, with correlation function  $R(\tau)$ . In this case, we can substitute the sample average with ensemble average to get

$$\begin{aligned}
 MSD &= E_x[x_k^2] + \frac{1}{n} \sum_{d=1}^{\infty} \left(\frac{n-1}{n}\right)^{(d-1)} E_x[x_{k+n-d}^2] - \frac{2}{n} \sum_{d=1}^{\infty} \left(\frac{n-1}{n}\right)^{(d-1)} E_x[x_k x_{k+n-d}] \\
 &= 2 \left[ R(0) - \frac{1}{n-1} \sum_{d=1}^{\infty} \left(\frac{n-1}{n}\right)^d R(n-d) \right]
 \end{aligned}$$

*Case II.*  $x(t) = \cos(\omega_0 t + \phi)$ , where  $\phi$  is uniformly distributed in  $[-\pi, \pi]$ . Note that averaging over  $\phi$  has a similar effect as sample averaging.

$$E_d A_k = \cos^2(bk + \phi) + E_d \{\cos^2[b(k+n-d) + \phi]\} - 2 \cos(bk + \phi) E_d \{\cos[b(k+n-d) + \phi]\}$$

Taking expectation over  $\phi$ , we get

$$E_d A_k = 1 - E_d \{\cos[b(n-d)]\}$$

independent of  $k$ . Therefore,

$$MDS = 1 - \frac{1}{n} \sum_{d=1}^{\infty} \left( \frac{n-1}{n} \right)^{(d-1)} \cos [b(n-d)]$$

or

$$MSD = 1 - \frac{2n \sin\left(\frac{b}{2}\right) \sin\left[b\left(n - \frac{1}{2}\right)\right] + \cos(bn)}{2n(n-1)[1 - \cos(b)] - 1}$$

Figure 13 shows  $MSD$  as a function of  $f_0$  for  $n = 4, 8, 16, 32$ . As can be expected, the scrambling is poor for low frequencies (at  $f_0 = 0$  there is no scrambling, no matter how large is  $n$ ). The cutoff frequency, where the scrambling starts to be effective, is a decreasing function of  $n$ . This means that the speech input should be high-passed before scrambling, the cutoff frequency depending on  $n$ .

## 6. Operation of the Decoder

We have seen that, in the coder, each sample is delayed  $n/2W$  seconds, on the average. Since the output of the decoder must be the original input sequence  $\{x_k\}$ , there has to be an additional delay in the decoder. Let the total delay be  $L$  samples (i.e.,  $L/2W$  seconds).

At time  $k/2W$ ,  $x_k$  is stored at cell  $a_k$ . If the total delay is  $L$  samples,  $x_k$  must be transmitted not later than  $(k + L - 1)/2W$  and, therefore, we must have  $a_{k+j} = a_k$  for some  $1 \leq j \leq L - 1$ . Hence, every integer  $1, \dots, n$  must appear at each subsequence of the key, of length

$L - 1$ . By similar arguments, we can show that the receiver storage must contain  $L$  cells.

It is clear that a Bernoulli sequence of index  $n$  violates the above condition for any  $L$ , since  $p(d \geq L + 1) = (n - 1/n)^L > 0$ . Therefore, the above analysis of the scrambling is only an approximation of the performance of an actual system. However, for large  $n$ , the analysis is asymptotically correct.

## 7. Conclusions

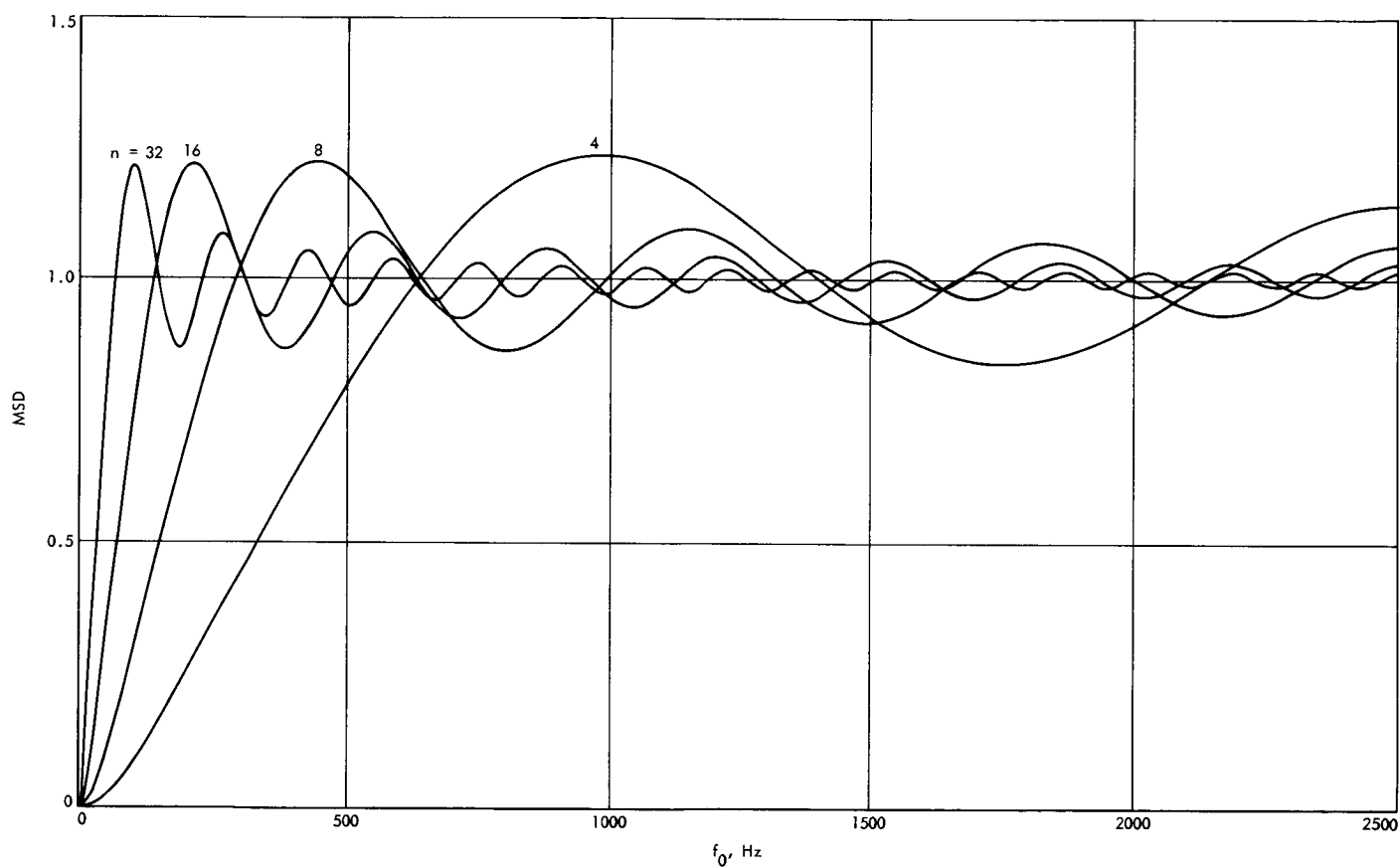
The results of this work quantify what we knew intuitively, i.e., that the increasing  $n$  improves the scrambling. We have shown that most of the energy is equally distributed in the bandwidth  $[-W, W]$ , and only a fraction  $\epsilon(n)$  remains around the input frequencies. Equivalently, the output is a signal with energy  $\epsilon(n)$  and a white noise with energy  $1 - \epsilon(n)$ . To determine  $n$ , we must know how small

$$\frac{\epsilon(n)}{1 - \epsilon(n)}$$

has to be to disguise the speech satisfactorily. This can be done experimentally.

## References

1. Wozercraft, J. M., and Jacobs, I. M., *Principles of Communication Engineering*. John Wiley & Sons, Inc., New York, N. Y., 1965.
2. Flanagan, J. L., *Speech Analysis Synthesis and Perception*. Academic Press, Inc., New York, N. Y., 1965.
3. Jolley, L. B. W., *Summation of Series*. Dover Publications, Inc., New York, N. Y., 1961.



**Fig. 13. Mean-square differences for  $n = 4, 8, 16, 32$**

## V. Communications Elements Research

### TELECOMMUNICATIONS DIVISION

#### A. RF Techniques Research: System Studies for Frequencies Above S-Band for Space Communications, R. Clauss

##### 1. Introduction

Efforts to develop low-noise receiving equipment for higher frequencies have been concentrated on maser design and materials at frequencies between 7 and 18 GHz. Improved maser performance may be obtained at the  $\theta = 90$  deg orientation. In the past, operational masers using ruby above 7 GHz have been operated at  $\theta = 54.7$  deg using the push-pull pumping method.

The use of multiple pumping at the  $\theta = 90$  deg orientation simplifies crystal orientation problems, improves dynamic stability, provides higher gain per unit length and reduces the maser noise temperature.

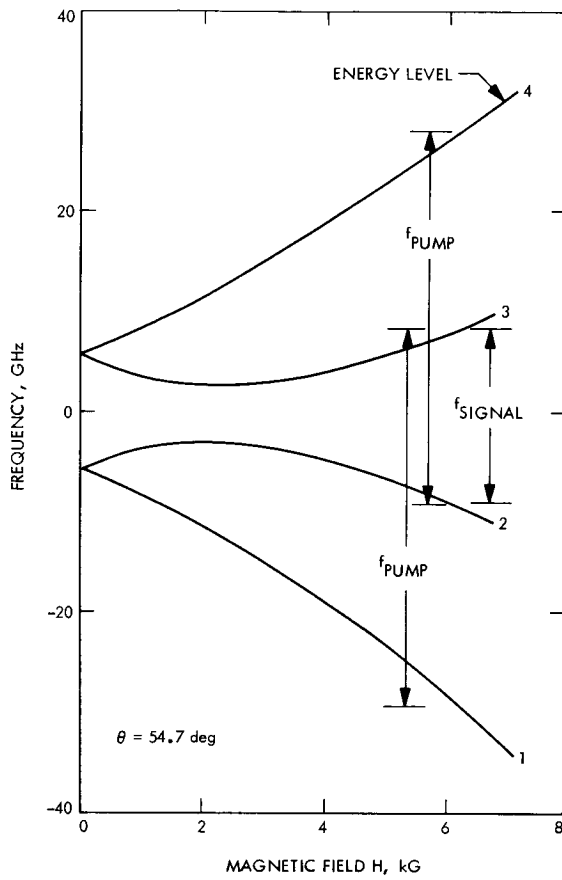
##### 2. Pumping Methods

Many pumping schemes for operation of solid-state masers were suggested between 1957 and 1961. Two methods of particular interest are described here.

Meyer (Ref. 1) reported operation of the push-push method using  $K_3Cr(CN)_6$  in October 1958.

Push-pull pumping using ruby was reported by Makhov, et al., (Ref. 2) in February 1958. Many masers operating at frequencies above 7 GHz have used the push-pull method. Push-pull pumping with a single-frequency pump source is possible in ruby at the orientation  $\theta = 54.7$  deg. The energy levels are shown in Fig. 1. A push-pull maser was built for use at 8450 MHz at JPL and is described in SPS 37-42, Vol. III, pp. 42-46.

A characteristic of the push-pull technique which causes degraded performance in operational situations is sensitivity of signal and pump frequency transitions to orientation change. The change of transition frequencies for both push-pull and push-push methods, with orientation, is shown in Fig. 2. The  $\theta = 90$  deg orientation tolerates misorientations as large as 1 deg without performance degradation (ten times less sensitive than the 54.7 deg orientation). This simplifies orientation and assembly procedures, and crystal quality requirements are reduced to within the present state-of-the-art growth capability for large crystals.



**Fig. 1. Energy levels for ruby at  $\theta = 54.7$  deg showing push-pull pumping levels**

The energy level diagram for  $\theta = 90$  deg is shown in Fig. 3. Recent investigations have used the push-push pump method with two separate pump sources. Advantages achieved with this combination are larger transition probabilities and higher measured inversion ratios than the push-pull method. Higher gain per unit length and lower noise temperature result. Reduced orientation sensitivity improves the dynamic stability of an operational maser.

### 3. Inversion Ratios

Measurements of inversion ratio have been made at frequencies between 7 and 18 GHz. Five pumping methods were used. Inversion ratios were measured at 4.5 and 1.8°K with push-push pumping. All other tests were made at 4.5°K. The results are shown in Table 1.

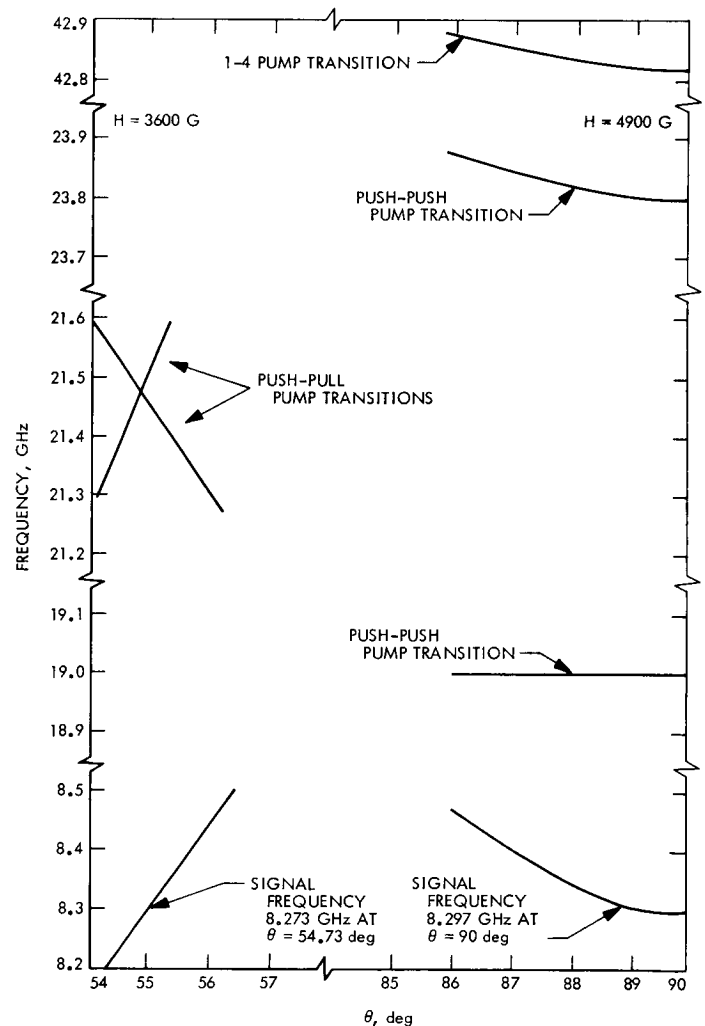
Siegman (Ref. 3) summarizes inversion ratio calculations for single and multiple pumped methods. Maximum ratios are achieved only when relaxation rates are optimum and sufficient power can be used to saturate the

pump transition. These maximum ratios are tabulated in Table 1 for comparison with measured values.

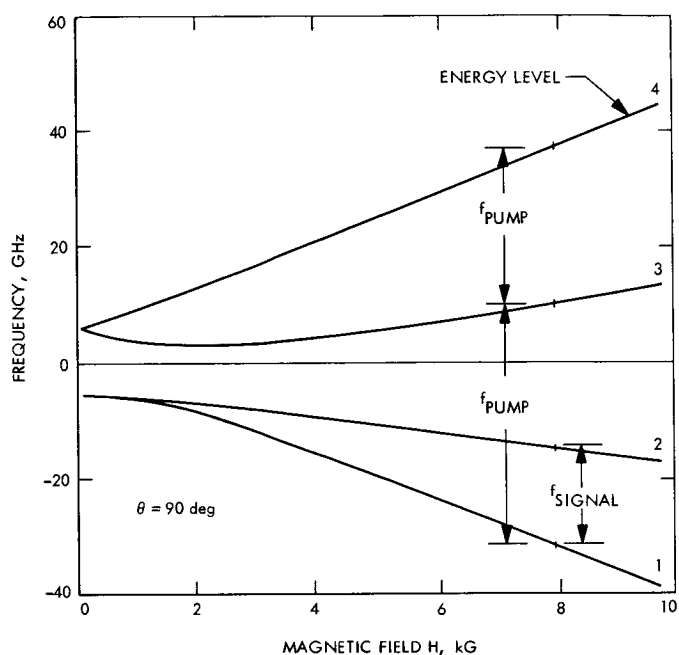
### 4. Orientation Merit Comparison

A figure of merit  $\Delta f_L / Q_m$  for maser material evaluation has been described in SPS 37-56, Vol. II, pp. 101-104. The linewidth of ruby used has been constant (approximately 60 MHz). The magnetic  $Q (Q_m)$  varies with frequency and orientation. The curves obtained for the figure of merit versus frequency in the two cases of interest are derived from Siegman (Ref. 3, p. 255):

$$\frac{1}{Q_m} \approx 10^{-18} \frac{hf}{kT} \frac{N}{n} \frac{IM^2F}{\Delta f_L}$$



**Fig. 2. Change in signal and pump frequencies with orientation changes for ruby**



**Fig. 3. Energy levels for ruby at  $\theta = 90$  deg showing push-push pumping levels**

where

$h$  = Planck's constant

$f$  = frequency, Hz

$k$  = Boltzmann's constant

$T$  = temperature, °K

$N$  = spin density, spins/cm<sup>3</sup>

$n$  = number of energy levels

$I$  = inversion ratio

$M^2$  = maximum value of the transition probability

$F$  = fill factor

$\Delta f_L$  = material line width, MHz

Ideal inversion ratios are assumed for the curves in Fig. 4.

A 0.05% ruby dielectric resonator was used at 16.8 GHz for inversion ratio and gain measurements. The same resonator was used for  $\theta = 90$  deg and  $\theta = 54.7$  deg. The

**Table 1. Measured and ideal inversion ratios**

Orientation $\theta$ , deg	Pump method (Energy level <sup>a</sup> )	Signal frequency, GHz (Energy level)	Pump frequency, GHz	Ideal inversion ratio	Measured inversion ratio
54.7	Push-pull (1-3, 2-4)	8.5 (2-3)	21.9	4.2	2.5
90	Single frequency (1-4)	8.5 (1-2)	43.5	4.1	2.5
90	Push-push (1-3, 3-4)	8.5 (1-2)	24.2, 19.3	4.1	3.2
90	Single frequency (1-3)	8.5 (1-2)	24.2	1.8	0.4
90	Single frequency (3-4)	8.5 (1-2)	19.3	—	0.2
90	Push-push (1-3, 3-4)	13.6 (1-2)	33.9, 24.6	3.3	3.0
90	Push-push (1-3, 3-4)	14.2 (1-2)	35.2, 25.2	3.2	2.9 <sup>b</sup>
90	Single frequency (1-4)	16.8 (1-2)	68.3	3.1	-0.1 <sup>c</sup>
90	Single frequency (1-3)	16.8 (1-2)	40.3	1.4	0.1
90	Single frequency (3-4)	16.8 (1-2)	28.0	—	0.0
90	Push-push (1-3, 3-4)	16.8 (1-2)	40.3, 28.0	3.1	2.9
54.7	Push-pull (1-3, 2-4)	16.8 (2-3)	37.3	3.4	2.5

<sup>a</sup>See Figs. 1 and 3.

<sup>b</sup>At 1.8°K.

<sup>c</sup>Pump transition not saturated.

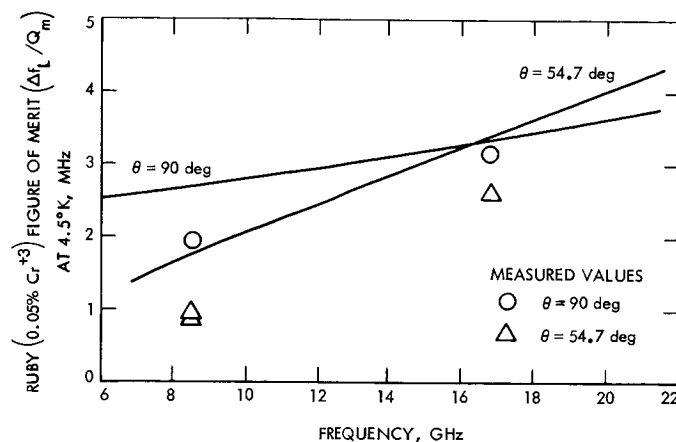


Fig. 4. Figure of merit versus frequency for ruby at  $\theta = 90$  deg and  $\theta = 54.7$  deg

filling factor for each case was considered in the calculation of  $Q_m$ . The difference between the figure of merit obtained with measured values and the predicted curves is explained by lower than ideal inversion ratios (Table 1).

Test data at 8.5 GHz were obtained using comb-type traveling-wave structures and ruby bars individually oriented for each case.

Measured inversion ratios for the push-push method exceeded those measured for the push-pull method at all frequencies tested. This result indicates a superior figure of merit for the  $\theta = 90$  deg orientation at frequencies below 20 GHz.

#### References

1. Meyer, J. W., "The Three Level Maser," Report 37-32, Lincoln Laboratory, Massachusetts Institute of Technology, Oct. 21, 1958.
2. Makhov, G., et al., "Maser Action in Ruby," *Phys. Rev.*, Vol. 109, No. 4, pp. 1399-1400, Feb. 15, 1958.
3. Siegman, A. E., *Microwave Solid State Masers*, pp. 255, 292. McGraw-Hill Book Co., Inc., New York, 1964.

### B. RF Techniques Research: S/X-Band Experiment, G. Levy, R. Dickinson, and C. Stelzried

#### 1. Introduction

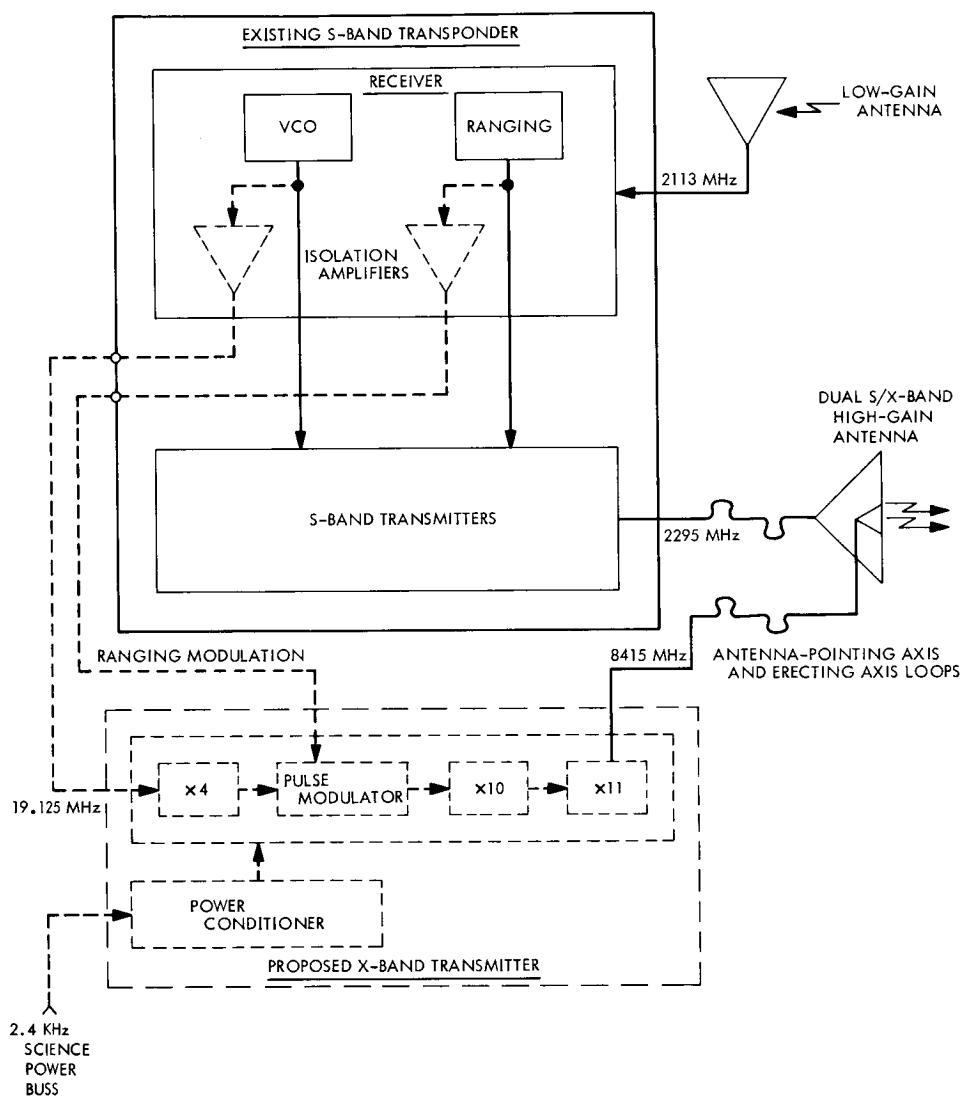
Dispersive measurements have been used to measure the charged particle environment between the spacecraft and the ground in many programs, such as the *Pioneer* series (Ref. 1). At the same time there is a need, on the basis of communication system requirements, to go higher

in frequency, particularly to X-band (Ref. 2). A desirable way to meet both these requirements and also obtain flight experience at X-band is to perform a dispersive measurement using the *Viking* orbiter as a platform for transmitting coherent range modulated S- and X-band radiation. A proposal was submitted to the *Viking* Project and was accepted to perform an S/X-band scientific experiment on the orbiter.

#### 2. Discussion

The S/X-band spacecraft equipment will consist of the standard S-band transponder with some minor modifications. The modifications will include isolation amplifiers and output jacks for the range modulation and the voltage-control oscillator (VCO). The VCO output will be multiplied in a separate multiplier chain which will produce a nominal 200 MW of range modulation energy at 8415 MHz (Fig. 5). The range modulation for both S- and X-band signals will be identical, and the carriers of the two signals will be phase coherent. As the S- and X-band signals propagate from the spacecraft to the earth, they will pass through the charged-particle medium. This charged-particle environment will produce both a dispersive phase acceleration and group retardation. By receiving these signals with a special configuration at the Mars Deep Space Station (DSS) it will be possible to measure the group delays and phase acceleration. From these data one can tell not only the total integrated electron content but also its time rate of change.

At the Mars DSS, there will be a special configuration for receiving both frequency signals simultaneously. This configuration is illustrated in Fig. 6. It will consist of a special S/X-band feed horn which will separate the two



**Fig. 5. Proposed spacecraft S/X-band transmitter block diagram**

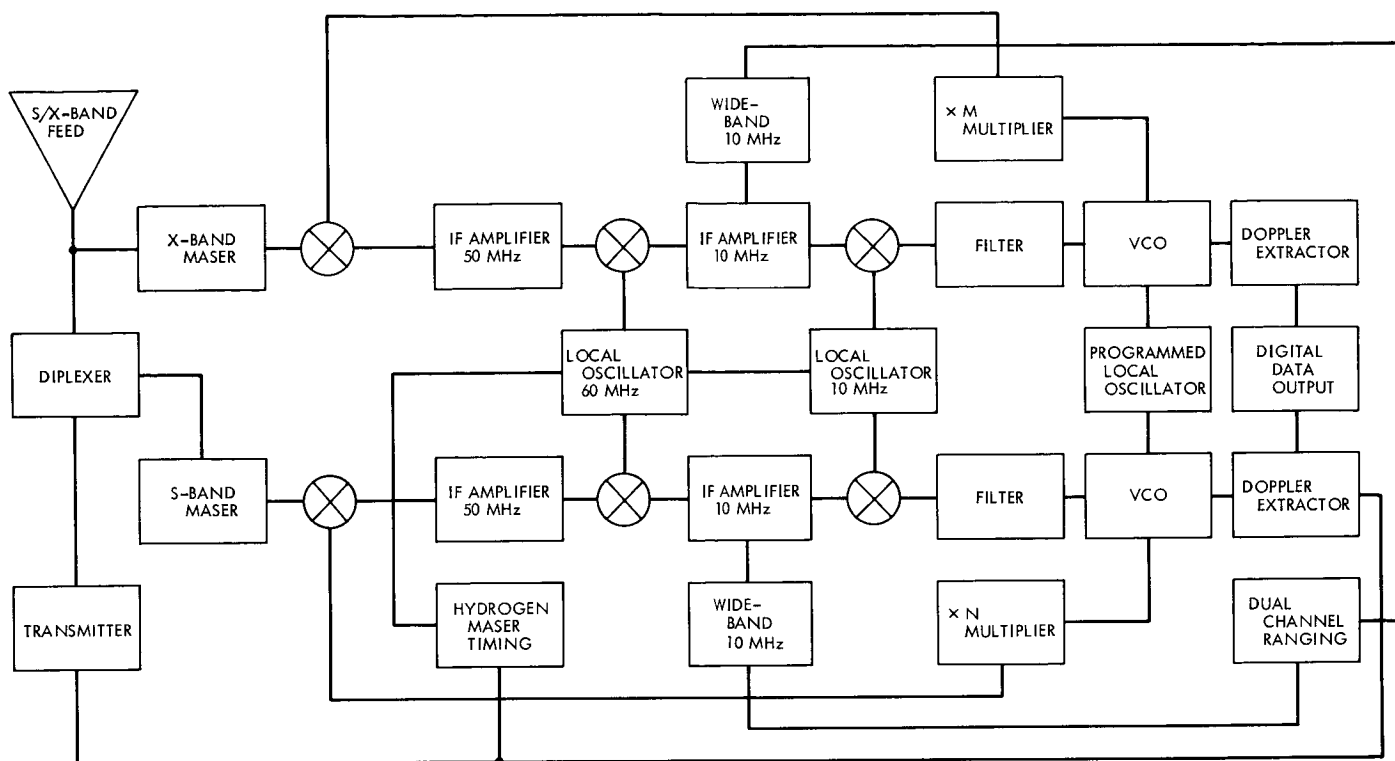


Fig. 6. Proposed S/X-band ground receiver block diagram

frequencies. The X-band signal will go to a maser, and the S-band signal will go to a diplexer and then to a maser. From the masers, the signals will go through parallel channels of the Block IV receiver. Separate doppler extractors will be used to monitor the change in phase of the two carrier frequencies and thereby measure the phase dispersion. A dual channel modification of the ranging machine will be used to measure the differential time delays between S and X-band to obtain an absolute measurement of the integrated electron density. The spacecraft components are now being breadboarded. The Block IV receiver is scheduled for initial installation at the Venus DSS in July, 1970.

Phase and group delay stability in both the spacecraft transponder and ground station will determine the accuracy and resolution obtainable by these dispersive measurements.

#### References

1. Koehler, R. L., "Radio Propagation Measurements of Pulsed Plasma Streams From the Sun Using *Pioneer* Spacecraft," *J. Geophys. Res. Space Phys.*, Vol. 73, No. 15, pp. 4883-4894, Aug. 1, 1968.

2. Potter, P. D., et al., *A Study of Weather-Dependent Data Links for Deep Space Applications*, Technical Report 32-1392, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 15, 1969.

### C. Quantum Electronics: Optical Communications Components: 10.6 $\mu$ m Heterodyne Radiometer, C. Finnie

#### 1. Radiometer Design

A 10.6- $\mu$ m heterodyne receiver was assembled for laboratory evaluation of photoconductor detector performance for infrared radiometry, communications, and geodesy applications.<sup>1</sup> The mixer is a compensated copper-doped germanium detector supplied by the Santa Barbara Research Center, Goleta, Calif. The detector is liquid helium-cooled. Its output signal is preamplified by a field effect transistor (FET) amplifier (20 k $\Omega$  input impedance) mounted on the 78°K stage of the detector dewar in close proximity to the detector crystal. The input to the radiometer was chopped at 130 Hz and synchronously detected with a Princeton Applied Research Model HR-8 Lock-in

<sup>1</sup>Much of the experimental work described in this article was performed by J. Siddoway.

amplifier. A  $100^\circ\text{K}$  increase above the  $300^\circ\text{K}$  (room temperature) chopper reference was detected with a signal-to-noise ratio of approximately 3 (Fig. 7).

A block diagram of the receiver is shown in Fig. 8. The local oscillator is a single-mode Sylvania Model 948  $\text{CO}_2$  laser. The circuit diagram of the cooled FET preamplifier is shown in Fig. 9. Carbon film resistors and Corning glass capacitors were used for the  $77^\circ\text{K}$  amplifier components. Figure 10 is a photograph showing the physical layout of the radiometer optical elements mounted on a granite table, and the electronic instrument rack.

## 2. Theoretical Performance

The limiting noise process for an ideal detector at  $10\text{ }\mu\text{m}$  is the quantized nature of photon arrival and detection.

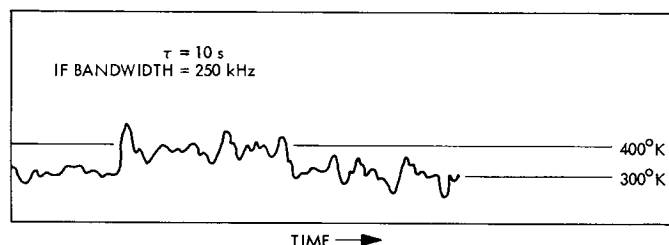


Fig. 7. Strip chart record of radiometer sensitivity

A figure of merit  $F_q$  has been defined for the quantum noise power contributed by an infrared receiver similar to the microwave noise figure where  $KTB$  is replaced by  $h\nu B$  (Ref. 1):

$$F_q = \frac{\text{measured noise power}}{\text{theoretical noise power}} = \frac{P_s}{(S/N) h\nu B}$$

where

$P_s$  = signal power

$S/N$  = measured signal-to-noise ratio

$F_q$  = quantum noise figure of merit

$h$  = Plank's constant

$\nu$  = frequency of the incident radiation

$B$  = bandwidth

$K$  = Boltzmann's constant

$T$  = absolute temperature

Therefore a minimum spectral density of

$$h\nu = -167.3 \text{ dBmW/Hz}$$

(1)

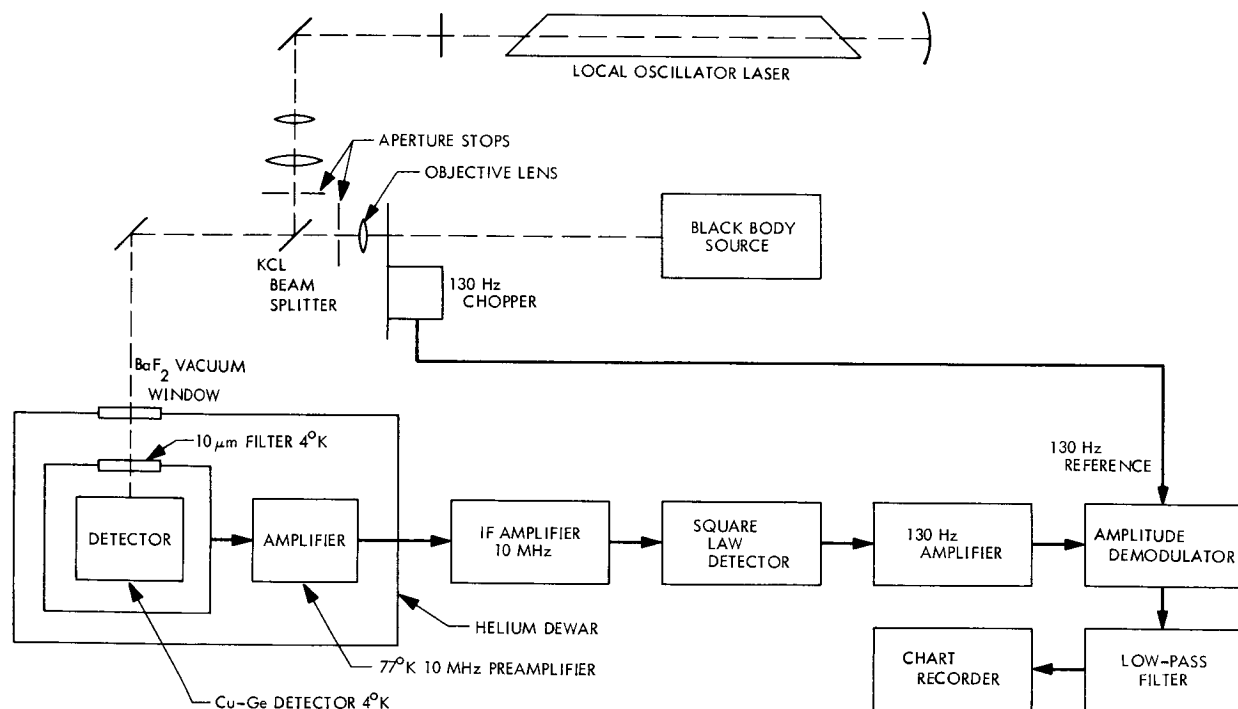
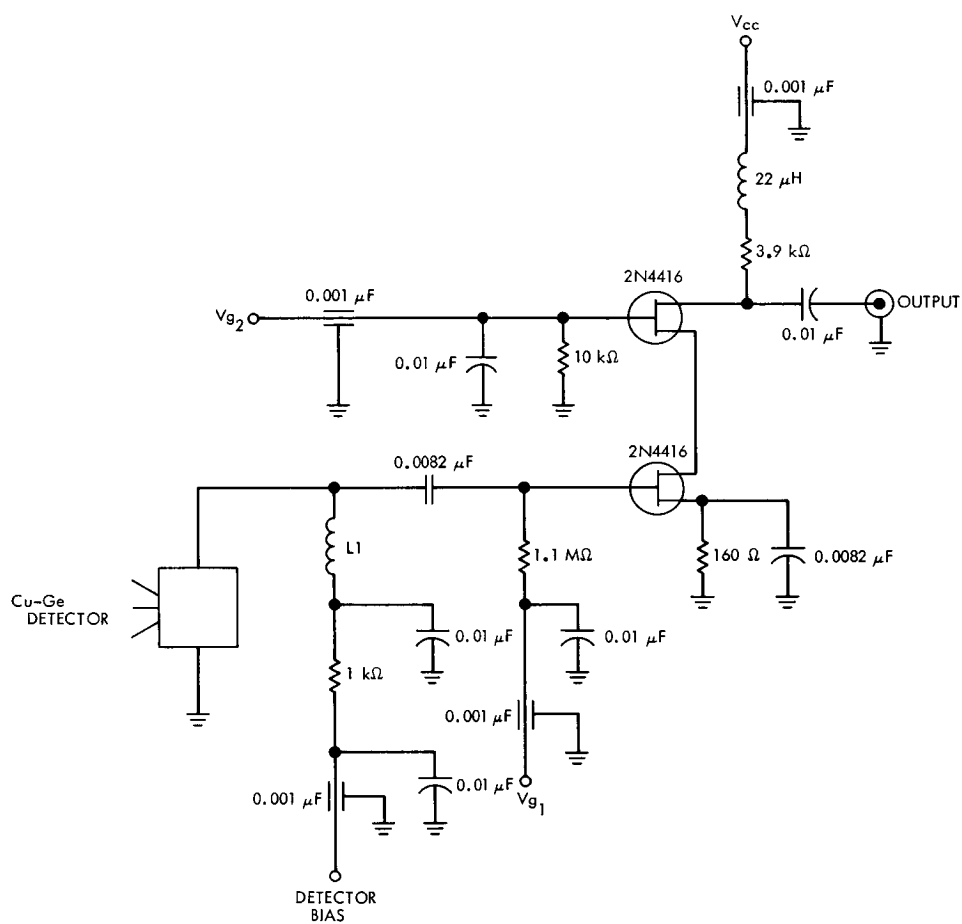


Fig. 8.  $10.6\text{-}\mu\text{m}$  heterodyne radiometer



L1 RESONATE WITH DETECTOR CAPACITY TO 10 MHz

**Fig. 9. 10-MHz preamplifier for 77°K ambient temperature**

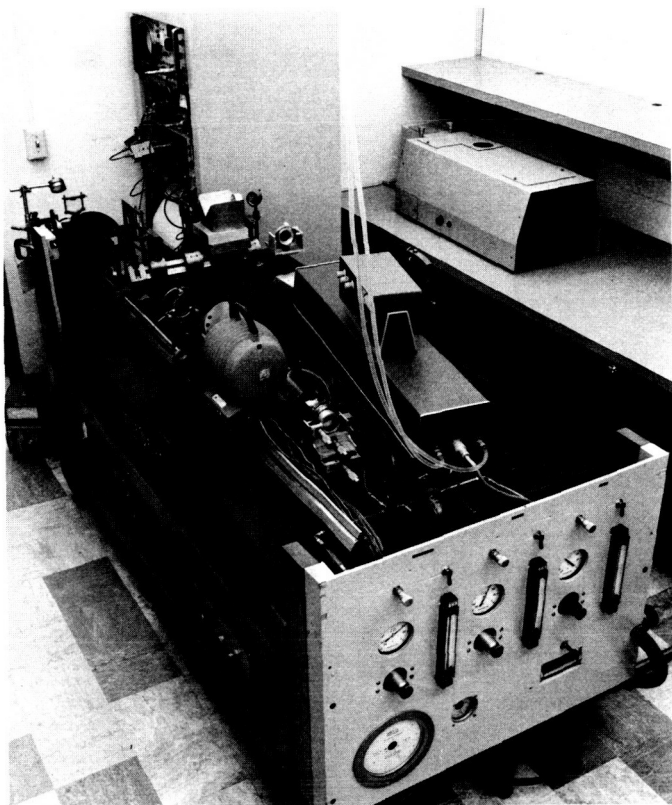


Fig. 10. Heterodyne receiver

is attainable at  $10.6 \mu\text{m}$ . An additional 5.5 dB can be attributed to the detector type (photoconductor) and detector quantum efficiency, decreasing the minimum detectable sensitivity to  $-161.8 \text{ dBmW/Hz}$ .

A limitation imposed in infrared heterodyne radiometer systems not present in microwave design is a detector tens of wavelengths in diameter. The large relative size limits the maximum solid angle from which the detector can accept coherent radiation (Ref. 2).

$$\frac{A_s A_L}{L^2} \approx \lambda^2 \quad \text{or} \quad \omega_L \approx \frac{\lambda^2}{A_s} \quad (2)$$

$\omega_L$  = solid angle at the detector subtended by the telescope aperture

$L$  = telescope focal length

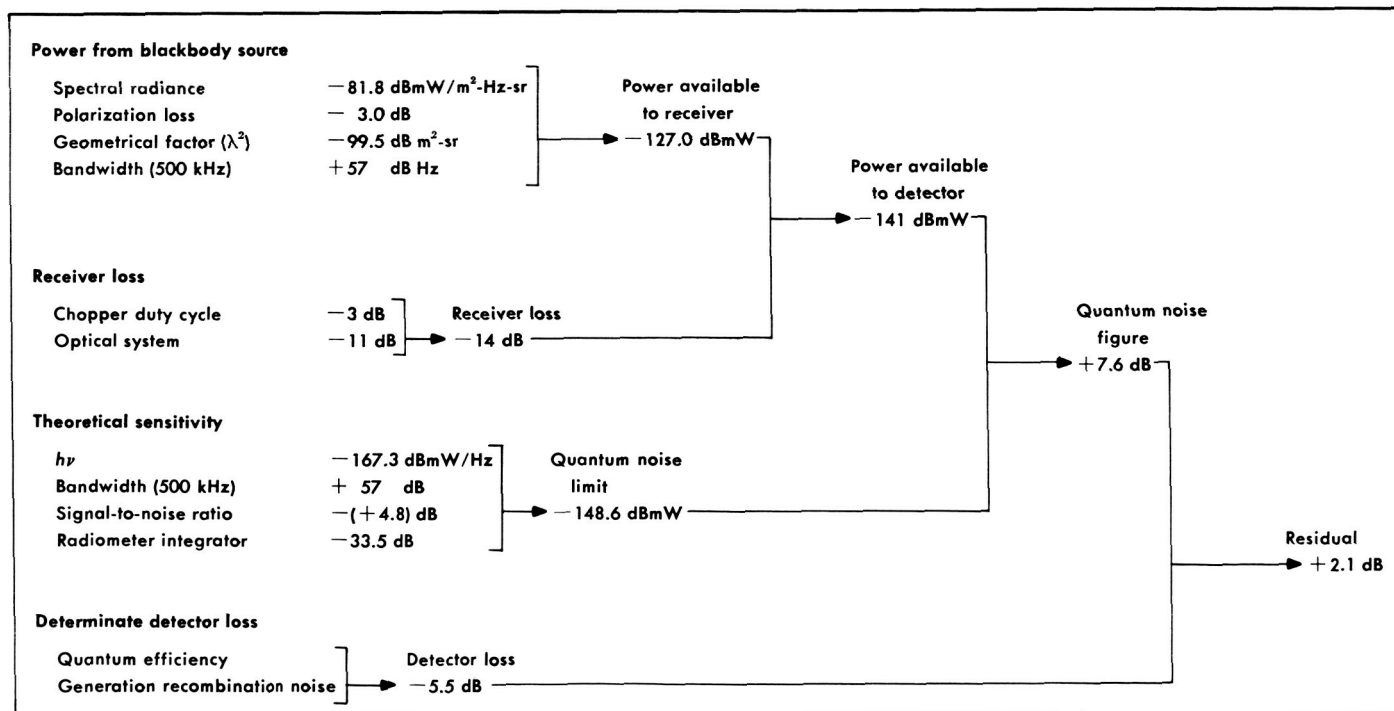
$A_s$  = the detector area

$A_L$  = input aperture area

$\lambda$  = wavelength of the received energy

From Eq. (2), the diffraction limited spot area is directly proportional to the telescope  $f$ -number squared. At  $10.6 \mu\text{m}$ , the  $f$ -number must be approximately 150 to

Table 2. Detector and radiometer performance evaluations



keep the spatial phase uniform over the 1-mm detector diameter.

### 3. Performance Evaluation

Table 2 itemizes contributions to the system sensitivity. A positive sign on a table value defines a system degradation or an increase in noise power. For example,  $-4.8$  dB signal to noise ratio indicates a signal to noise ratio of 3. The residual 2.1 dB between the theoretical and expected mixer sensitivity is within present experimental accuracy. An improvement of overall system performance of approximately 6 dB by antireflection coating optical components should be possible. Focal ratios as small as  $f = 16$  were used independently and simultaneously on local oscillator and input apertures. Focal ratios less than  $f \approx 150$  were found to have no affect on the radiometer sensitivity. Present efforts are directed toward improving local oscillator stability, as this is now the limiting factor in the frequency resolution of the radiometer.

### References

1. Arams, F. R., Sard, E. W., Peyton, B. J., and Pace, F. P., "Infrared 10.6 Micron Heterodyne Detection With Gigahertz IF Capability," *IEEE J. Quant. Electron.*, Vol. QE-3, No. 11, Nov. 1967.
2. Siegman, A. E., "The Antenna Properties of Optical Heterodyne Receivers," *Proc. IEEE*, Vol. 54, No. 10, October 1966.
3. Kruse, P. W., McGlauchlin, L. D., and McQuistan, R. B., *Elements of Infrared Technology, Generation, Transmission and Detection*, John Wiley & Sons, Inc., New York, 1962.

## D. Spacecraft Antenna Research: An RF Study of Reflector Surface Materials for Spacecraft Antennas, K. Woo and T. Y. Otoshi

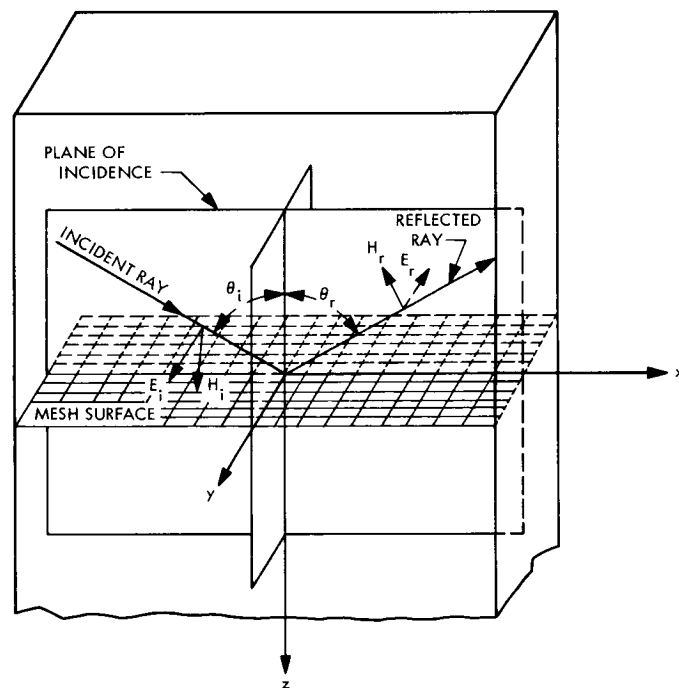
### 1. Introduction

For high RF performance, reflector antennas must have good surface reflectivity. In unfurlable antennas, the reflector surface is made of flexible metallic mesh. The RF reflectivity of such mesh is, therefore, of concern to antenna designers. For the Thermoelectric Outer Planet Spacecraft (TOPS) program, the use of a 14-ft-diam unfurlable antenna has been proposed. In supporting this program, a waveguide measurement technique capable of evaluating RF reflectivity of metallic mesh materials, as well as other reflecting surfaces, has been developed. The technique provides a reflectivity loss measurement accuracy of  $\pm 0.005$  dB. This article describes the theory and setup of this technique and also presents the measured RF reflectivity of various mesh and other reflecting materials.

### 2. Waveguide Measurement Technique

In the RF testing of a mesh material for unfurlable antenna applications, the accuracy of reflectivity measurements is generally considered to be of primary importance. Best accuracy in measurements of reflection coefficients can usually be obtained through the use of waveguide methods.

A waveguide technique was used previously at JPL for investigating the relationship between mesh porosity and RF leakage (SPS 37-19, Vol. III, pp. 46-48 and SPS 37-20, Vol. IV, pp. 135-137). As was pointed out in the referenced SPS articles, the waveguide method requires that the mesh sample has a periodic structure. The free-space situation being simulated by the waveguide method is an infinite two-dimensional array of the mesh and a plane wave which is incident upon the array at an oblique angle of incidence  $\theta_i$ . The equivalent relationships between the free space and waveguide test configurations can be seen in Fig. 11. When the waveguide supports only the  $TE_{10}$



$E_i, H_i$  = INCIDENT WAVE ELECTRIC AND MAGNETIC FIELD VECTORS  
 $E_r, H_r$  = REFLECTED WAVE ELECTRIC AND MAGNETIC FIELD VECTORS  
 $\theta_i, \theta_r$  = ANGLE OF INCIDENCE AND ANGLE OF REFLECTION

**Fig. 11. Free-space test configuration being simulated in a  $TE_{10}$  mode waveguide**

mode, the angle of incidence is determined by the waveguide cut-off frequency  $f_c$  and the operating frequency  $f$  by the relationship

$$\theta_i = \sin^{-1} \left( \frac{f_c}{f} \right) \quad (1)$$

For an operating frequency of 8448 MHz in WR 112 waveguide, the angle of incidence is 38.5 deg.

Simulation of the above described free space situation is accurate: (1) if the mesh pattern is imaged to waveguide walls, or (2) if there are many basic cells of the periodic structure filling the waveguide cross-sectional area. The portions of incident power reflected by the mesh, absorbed by the mesh, and transmitted through the mesh are related to waveguide measured values of return loss and insertion loss by the equations

$$\text{Reflected power ratio} = P_r/P_i = |\Gamma_s|^2 10^{-RL/10} \quad (2)$$

Absorbed power ratio =

$$P_a/P_i = 1 - |\Gamma_s|^2 10^{-RL/10} - 10^{-IL/10} \quad (3)$$

$$\text{Transmitted power ratio} = P_t/P_i = 10^{-IL/10} \quad (4)$$

where

$|\Gamma_s|$  = voltage reflection coefficient of the reference short-circuit used in the return loss measurement

$RL$  = return loss (or reflectivity loss) of the mesh as measured relative to that of the reference short-circuit, dB

$IL$  = insertion loss when the mesh is inserted into a nonreflecting waveguide system, dB

The advantages of using the waveguide technique over free space methods are:

- (1) Through the use of tuned reflectometer waveguide techniques, it is feasible to measure power reflection coefficients of the order of 0.1 dB to accuracies of  $\pm 0.005$  dB.
- (2) Measurements of transmission coefficients in waveguide do not suffer from possible errors due to edge diffraction or multipath.
- (3) In waveguide, it is easier to investigate the changes in reflectivity properties due to placing various

types of terminations behind the mesh sample. A severe degradation of reflectivity properties may occur if a short circuit is positioned at a critical distance behind the mesh material.

- (4) The sample size required for the waveguide method is generally smaller than that required by free space methods. It is, therefore, easier to detect flaws or nonuniformities because many test samples can be cut from the same mesh cloth.

Figure 12 shows a block diagram of the WR 112 waveguide system which was operated at 8448 MHz. When the detected outputs of couplers 1 and 2 were compared, the system operated as a dual-channel-tuned reflectometer. When the detected outputs of coupler 3 were compared before and after insertion of the test sample, the system operated as an insertion loss measurement system. The voltage reflection coefficients of the equivalent generator and load assemblies at the test reference planes were tuned to be negligibly small. These reflection coefficients are denoted in Fig. 12 by  $\Gamma_{zi}$  and  $\Gamma_L$ , respectively. It appears that the waveguide measurement system is capable of providing an overall accuracy of about  $\pm 0.005$  dB on the measurement of small reflectivity losses and an overall accuracy of  $\pm 0.1$  dB per 10 dB on the measurement of insertion losses.

Figure 13 shows a waveguide sample holder that was designed to hold the mesh material after it had been stretched. Through the use of this sample holder, the mesh pattern could be oriented at different angles with respect to the linearly polarized electric field. It should be pointed out that although the mesh pattern orientation can be changed, the waveguide method is limited to the case of a linearly polarized wave whose electric field points in a direction perpendicular to the plane of incidence (Fig. 11).

When the sample holder (containing a sample) is inserted into the waveguide measurement system, the mesh material is compressed between the flange face on the sample holder and the flange on a waveguide load assembly. To prevent crushing of the plating on the mesh material, the center area of the sample holder flange face was counterbored to a depth approximately equal to the minimum thickness of the mesh. The actual depth of step used was a compromise between good electrical contact and tolerable amount of crushing of the material. A series of preliminary RF tests were made to verify that the step chosen was optimum for the particular type of material being tested.

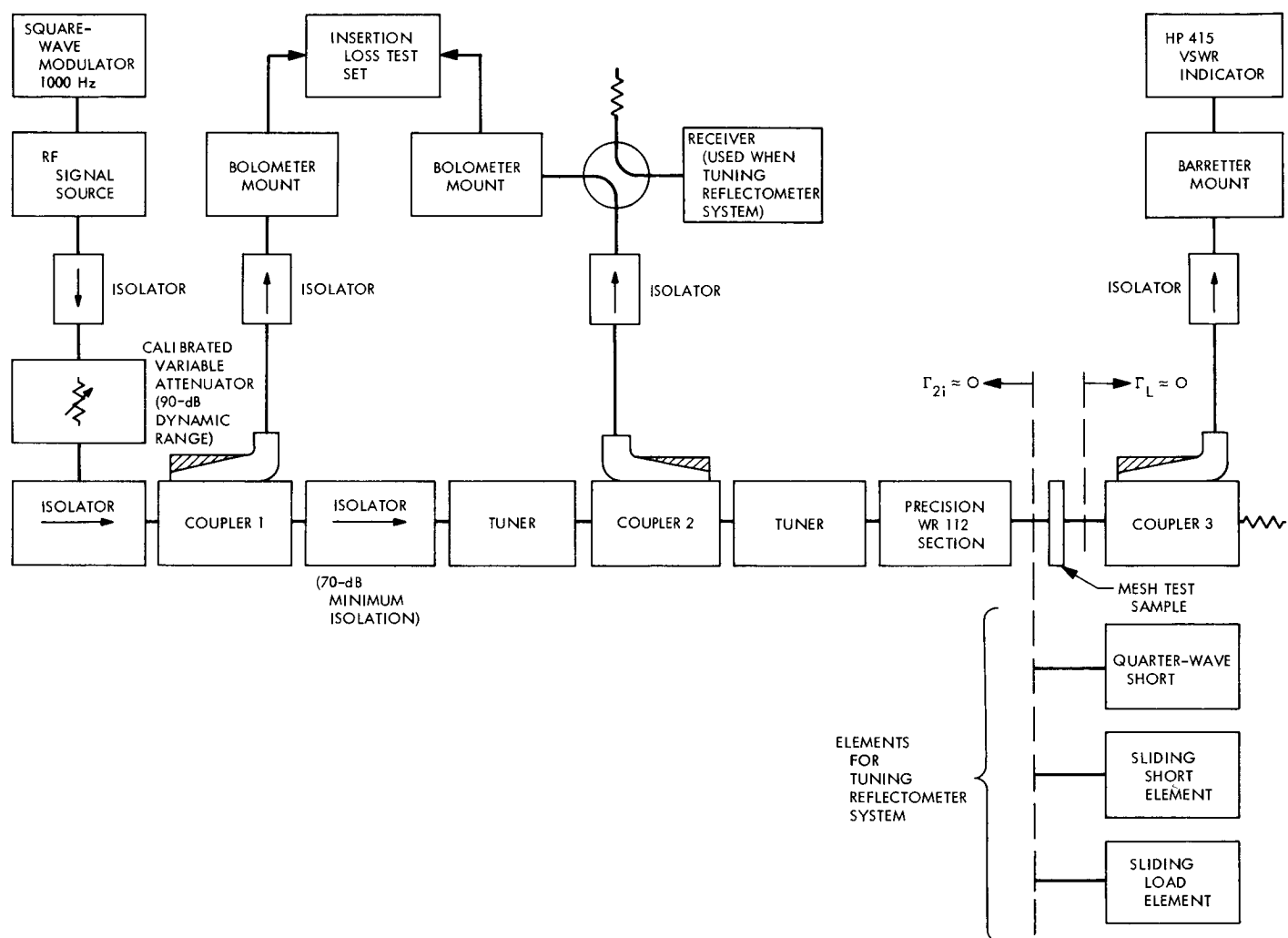
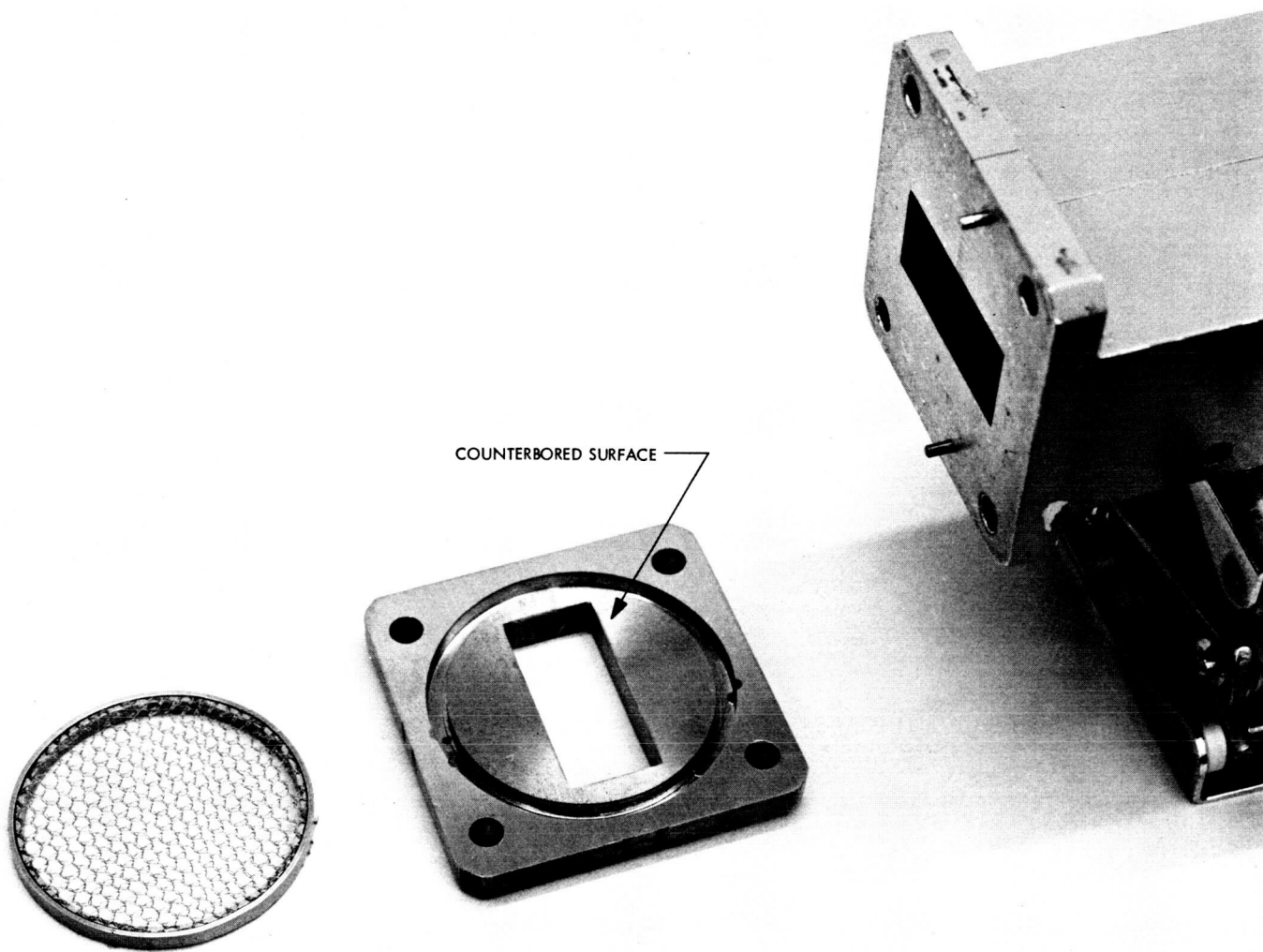


Fig. 12. Block diagram of waveguide system for mesh measurements



**Fig. 13. Waveguide mesh sample holder**

### 3. Results

By using the waveguide test setup and the sample holder, the RF reflectivity of two candidate mesh materials for TOPS, gold-plated Chromel-R and copper-coated Dacron, has been investigated. The gold-plated Chromel-R mesh (Fig. 14a) is composed of nichrome-type base material (74% Ni, 20% Cr, 3% Al, 3% Fe) plated with approximately 40  $\mu$ in. of nickel and then 5  $\mu$ in. of gold on the outside. The mesh is of tricot knit and is stretchable. The copper-coated Dacron mesh (Fig. 14b) is composed of Dacron strands (double twisted strands in the 0-deg direction and single strand in the 90-deg direction) coated with approximately 6  $\mu$ in. of copper and 90  $\mu$ in. of silicon protective layer on the outside. The mesh is not stretchable.

Tests made at 8448 MHz show that the RF reflectivity of both mesh materials varies from sample to sample and with mesh orientation. The RF reflectivity of the gold-plated Chromel-R also varies slightly with mesh tension. The variability of RF reflectivity loss for each mesh observed from all the samples tested in various orientations and tensions is shown in Table 3. As can be seen, the gold-plated Chromel-R mesh, in general, offers better RF reflectivity

than the copper-plated Dacron mesh. The inclusion of the data of nickel-plated (outer gold-plating stripped) and unplated Chromel-R is intended to show the effect on reflectivity in case of deterioration of the platings of a gold-plated Chromel-R mesh.

Table 4 shows the RF reflectivity of a typical sample of the gold-plated Chromel-R mesh for various mesh orientations and tensions. Table 5 shows the RF reflectivity of two samples of the copper-coated Dacron mesh for various mesh orientations. Sample A represents the typical performance of the majority of the copper-coated Dacron samples, and Sample B represents the best overall performance observed of all copper-coated Dacron samples tested. The inclusion of the insertion loss data is for the purpose of evaluating the dissipative characteristic of each mesh. As shown in Tables 4 and 5, neither mesh is isotropic.

All reflectivity loss data presented in this article are referenced to the reflection coefficient of a perfect reflector (i.e.,  $|\Gamma_s| = 1$ ). By applying Eqs. (2) through (4) to the data in Tables 4 and 5, the portions of incident power reflected by, absorbed by, and transmitted through each mesh can be determined. For example, for the case of gold-plated Chromel-R mesh with 45-deg orientation and medium tension, calculation shows that the power incident on the mesh is 97.3% reflected, 1.8% absorbed by the mesh, and 0.9% transmitted through the mesh. For the case of copper-coated Dacron mesh with 45-deg orientation, the power incident on Sample A is 91.4% reflected, 8.2% absorbed by the mesh, and 0.4% transmitted through, and the power incident on Sample B is 97.8% reflected, 2.1% absorbed by the mesh, and 0.1% transmitted through. In general, it seems that most of the power which is not reflected is dissipated in the mesh for both materials.

**Table 3. Variability of RF reflectivity loss of Chromel-R mesh and copper-coated Dacron mesh at 8448 MHz**

Type of mesh		Reflectivity loss, dB	
		Best case	Worst case
Chromel-R	Gold-plated	0.039	0.145
	Nickel-plated (outer gold-plating stripped)	0.131	0.851
	Unplated (base material)	0.431	3.36
Copper-coated Dacron		0.008	0.638

**Table 4. Loss characteristics of gold-plated Chromel-R mesh at 8448 MHz**

Mesh orientation*	Reflectivity loss, dB			Insertion loss, dB		
	No tension (loose)	Medium tension	Maximum tension	No tension (loose)	Medium tension	Maximum tension
0-deg direction parallel to E-field	0.095	0.091	0.091	24.3	23.1	21.9
45-deg direction parallel to E-field	0.126	0.119	0.115	21.3	20.6	20.3
90-deg direction parallel to E-field	0.080	0.080	0.079	25.4	24.1	23.1

\*For 0-, 45-, and 90-deg directions, see Fig. 14a.

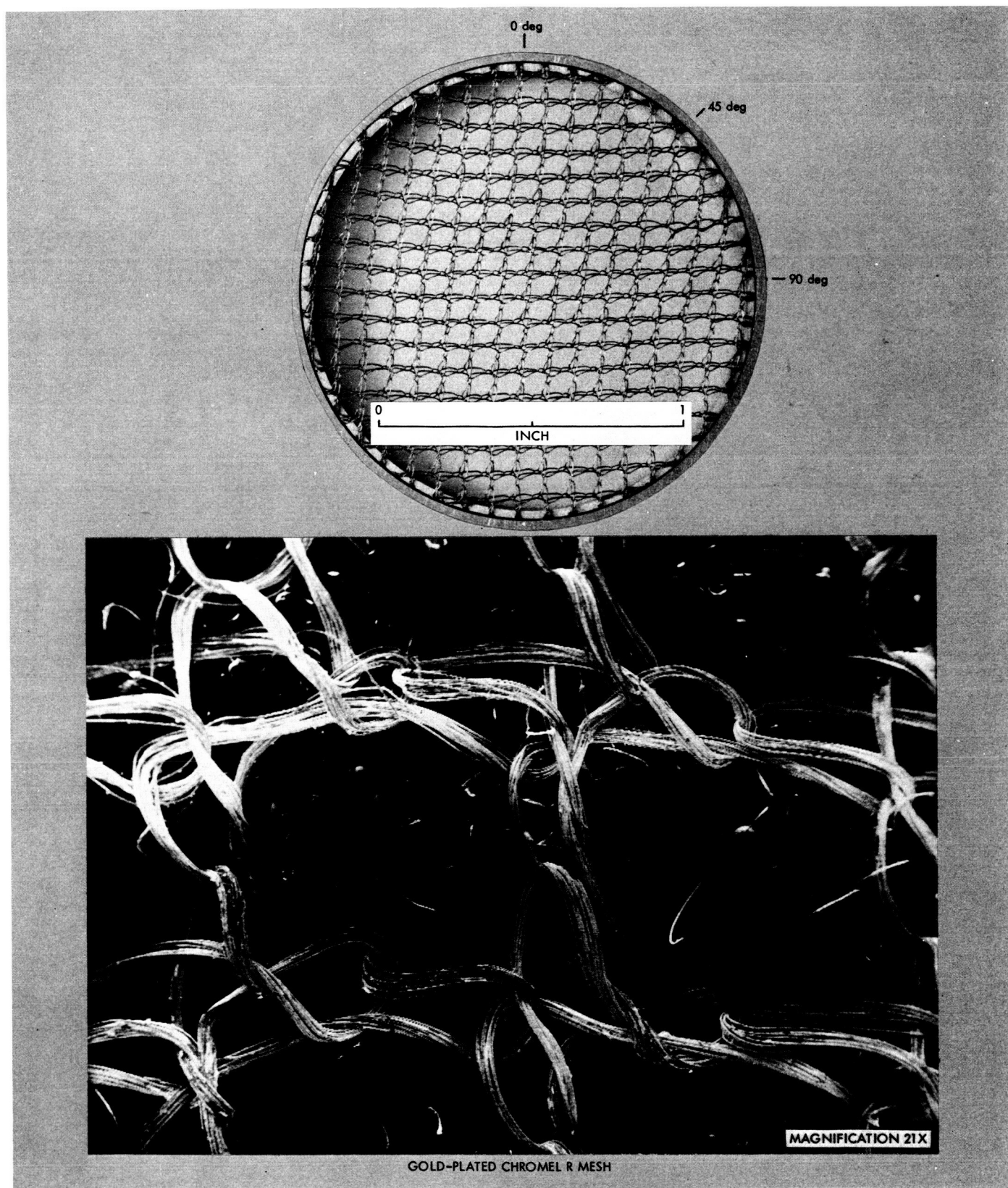


Fig. 14a. Configuration of gold-plated Chromel-R mesh

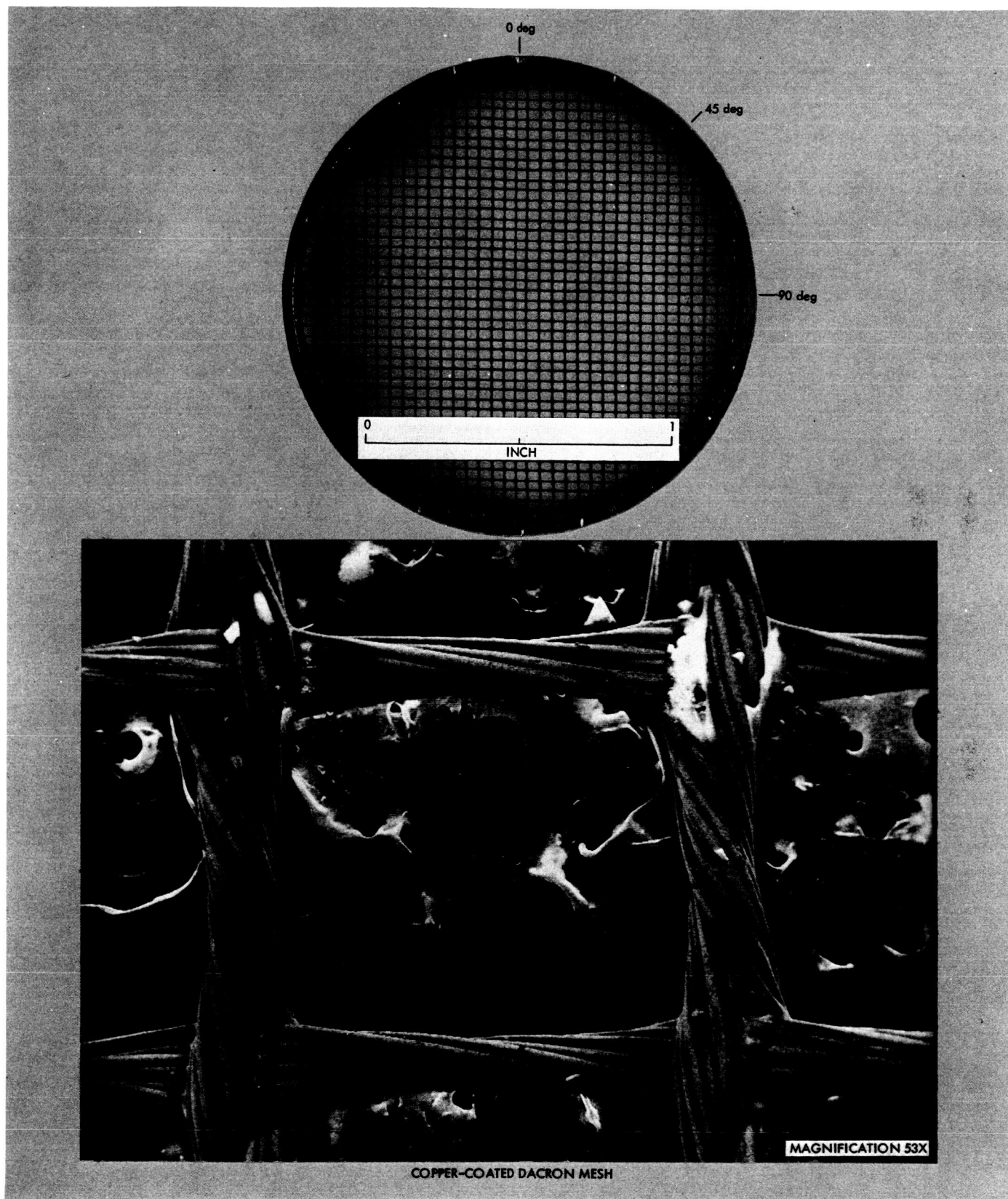


Fig. 14b. Configuration of copper-coated Dacron mesh

**Table 5. Loss characteristics of copper-coated Dacron mesh at 8448 MHz**

Mesh orientation <sup>a</sup>	Reflectivity loss, dB		Insertion loss, dB	
	Sample A	Sample B	Sample A	Sample B
0-deg direction parallel to E-field	0.550	0.074	25.1	28.1
45-deg direction parallel to E-field	0.390	0.096	23.8	29.0
90-deg direction parallel to E-field	0.560	0.166	22.8	28.2

<sup>a</sup>For 0-, 45-, and 90-deg directions, see Fig. 14b.

It should be noted that the data presented in Tables 3 through 5 are based on the samples as-received. It appeared that the gold-plating on the Chromel-R mesh was not in best condition and that the copper-coated Dacron mesh was fragile and sensitive to handling as far as RF reflectivity is concerned.

For possible use in other antenna applications, the RF reflectivity of aluminum and stainless steel sheets has also been investigated. At 8448 MHz, the RF reflectivity losses of aluminum and stainless steel sheets were found, respectively, to be 0.005 and 0.028 dB.

#### 4. Conclusions

A waveguide measurement technique capable of measuring the RF reflectivity of metallic mesh materials and other reflecting surfaces with an accuracy of  $\pm 0.005$  dB has been developed. Using the technique, the RF reflectivity of gold-plated Chromel-R and copper-coated Dacron mesh materials, aluminum and stainless steel sheets has been determined. It is found that, in general, the gold-plated Chromel-R mesh exhibits less RF reflectivity loss than that of the copper-plated Dacron mesh. The RF reflectivity losses of aluminum and stainless steel sheets are relatively small.

### E. Spacecraft Antenna Research: Radiation Characteristics of Mars Rough Lander Antennas, K. Woo

#### 1. Introduction

This article reports the radiation characteristics of the direct-link antennas of a Mars rough lander designed for the Capsule System Advanced Development (CSAD).<sup>2</sup>

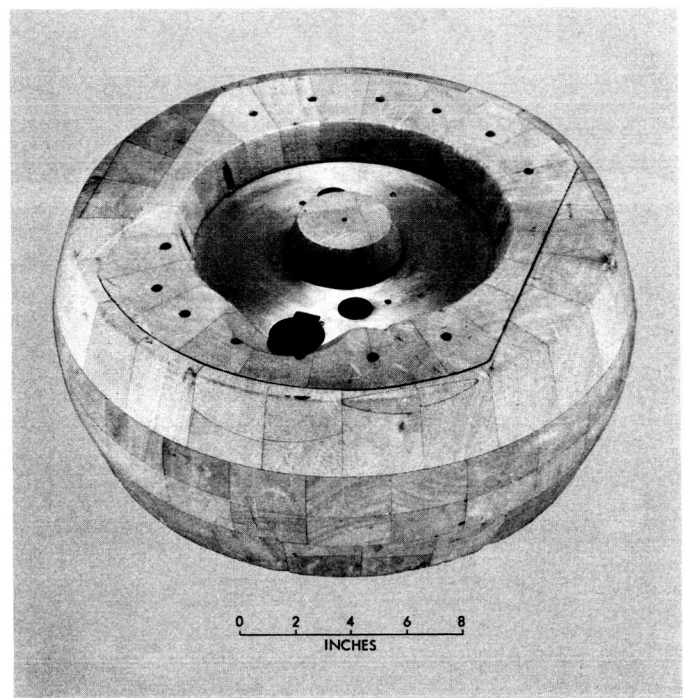
<sup>2</sup>The CSAD antenna design was completed in June 1968, but due to the unavailability of a suitable antenna range, the final patterns were not obtained until recently.

#### 2. Antenna Configuration

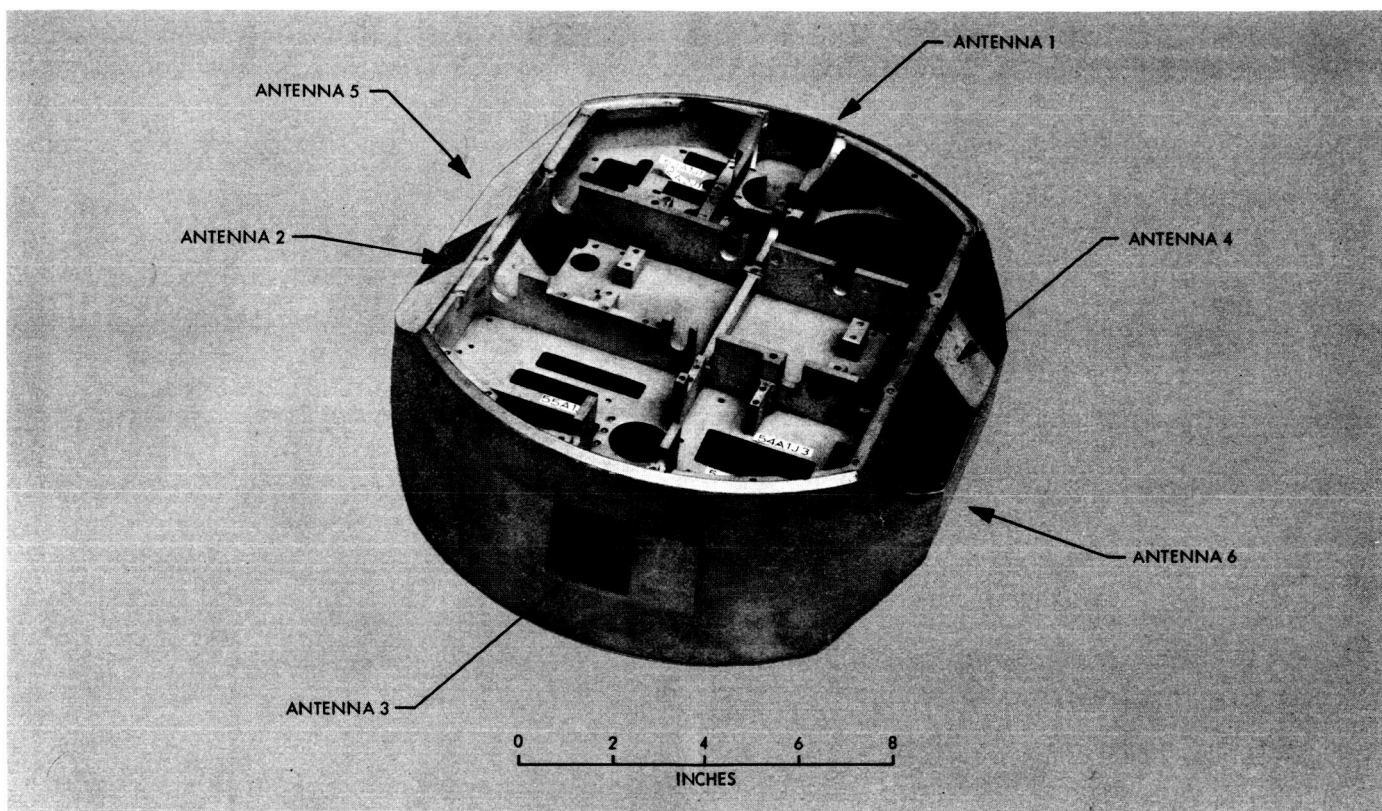
The CSAD Mars rough lander (Ref. 1) is shown in Fig. 15. The lander consists of six identical S-band (2298 MHz) circularly polarized direct-link antennas. The locations of the antennas is shown in Fig. 16 (a view of the lander chassis before the electronic components and the balsawood impact limiter were installed). Three of the antennas can be seen in this figure (foam-filled square regions); the other three are in similar positions on the other side of the lander. Each antenna is a square-cup radiator of the same design as reported in SPS 37-49, Vol. III, pp. 345-347. The internal construction of each antenna is shown in Fig. 17. For the purpose of supporting the feed probe and the cup walls to survive high impacts, each antenna cavity is filled with Eccofoam PT (density 27.5 lb./ft.<sup>3</sup>, yielding strength 870 psi, sterilizable). The balsawood impact limiter has a density of 7.9 lb./ft.<sup>3</sup>, is 3.0-in. thick above each antenna aperture and its vicinity, and is in grain with each antenna axis.

#### 3. Radiation Characteristics

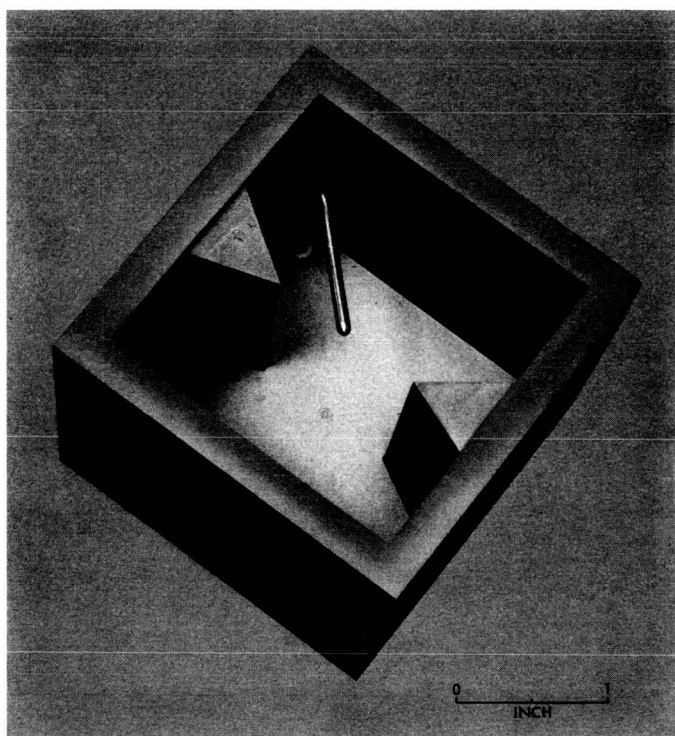
The lander of Fig. 15 was sterilized at 125°C (257°F) for about 60 h. Immediately after the sterilization, the radiation patterns were obtained. Only the patterns of the antennas 3 and 4 were taken, since the patterns of the remaining antennas are similar to either one or the other



**Fig. 15. CSAD Mars rough lander**



**Fig. 16. Locations of lander antennas**



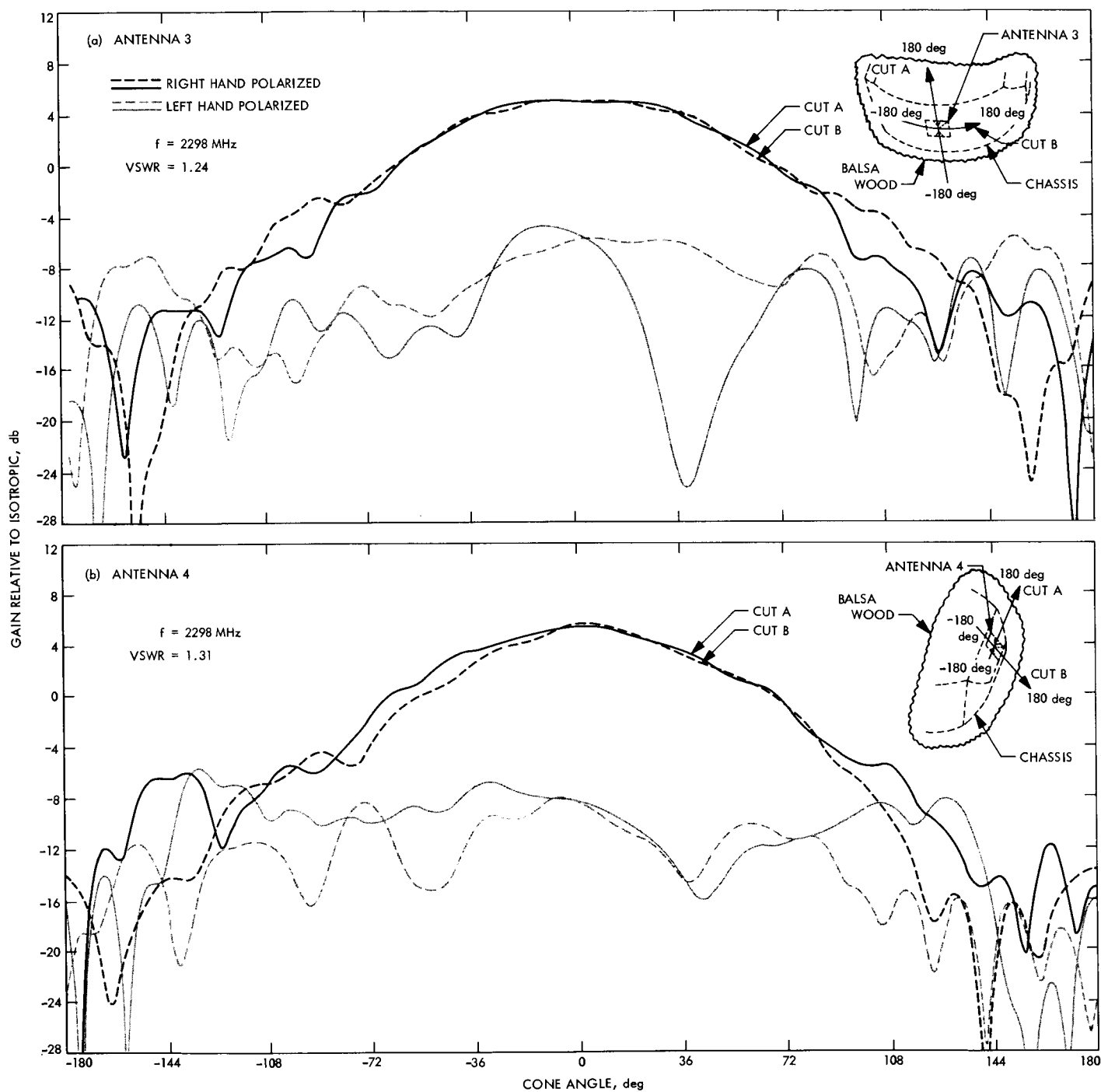
**Fig. 17. Internal construction of lander antennas**

of the two antennas due to the symmetry of the antenna locations.

Figure 18 shows the radiation patterns of two orthogonal cuts of each antenna. The patterns of both antennas are broader and have more ripples than the square cup antenna of the same design reported in SPS 37-49, Vol. III, due to the presence of the lander metal walls which act as ground planes. The patterns of antenna 3 have approximately equal beamwidths and are symmetrical with respect to the antenna boresight. The patterns of antenna 4 do not have equal beamwidths due to dissimilar lander wall configuration around the antenna. The ellipticity of each antenna is relatively high. Further improvement of ellipticity is desirable. The input voltage standing-wave ratios (VSWRs) of antennas 3 and 4 are 1.24 and 1.31, respectively. The antennas should survive an indirect impact of 10,000 g (as originally designed), although the maximum impact that the lander itself is designed for is only 2,500 g.

#### **4. Conclusions**

The performance of CSAD Mars rough lander antennas has been tested. It is found that the antennas provide good



**Fig. 18. Radiation characteristics of lander antennas: (a) Antenna 3, (b) Antenna 4**

patterns suitable for use as direct-link antennas. Further improvement in ellipticity and VSWR are desirable.

### Reference

1. Read, W. S., "Electronic Packaging and Cabling of a Mars Rough Lander," Paper ECP 10-00-01/1-69. Presented at the 1969 International Electronic Circuit Packaging Symposium at WESCON, San Francisco, Calif., August 1969.

## F. Spacecraft Antenna Research: Derivation of the Geometrical Optics Solution by the

**WKB Method, R. Woo**

### 1. Introduction

The fields produced by a dipole located in a cylindrical inhomogeneous plasma were obtained in SPS 37-60,

Vol. III, pp. 28-36, using the geometrical optics representation. The purpose of this article is to show that these results can also be derived by applying the WKB<sup>3</sup> approximation (Ref. 1) to the scalar wave equation in three dimensions. This fact is not obvious, since the vector wave equation for a three-dimensional inhomogeneous medium does not reduce to the scalar wave equation.

### 2. Discussion

The electric dipole with moment  $I\ell$  is located at  $(\rho_0, \phi_0, z_0)$ . It is assumed that the scalar field  $E$  satisfies the scalar wave equation

$$(\nabla^2 + k^2 n^2) E = -j\omega\mu I\ell \delta(\mathbf{r} - \mathbf{r}_0) \quad (1)$$

where  $k$  is the free space wave number,  $n$  is the refractive index and  $\mathbf{r}$  is the position vector. Using cylindrical coordinates Eq. (1) can be written

$$\left[ \frac{1}{\rho} \frac{\partial}{\partial \rho} \left( \rho \frac{\partial}{\partial \rho} \right) + \frac{\partial^2}{\partial z^2} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2} + k^2 n^2 \right] E(\rho, \phi, z) = -j\omega\mu I\ell \delta(z - z_0) \delta(\phi - \phi_0) \delta(\rho - \rho_0)/\rho \quad (2)$$

Taking the Fourier transform of Eq. (2) in  $z$  and then  $\phi$ , yields the following result:

$$E = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} dh \int_{-\infty}^{\infty} d\nu U \exp[-jh(z - z_0) - j\nu(\phi - \phi_0)] \quad (3)$$

where  $U$  satisfies

$$\left[ \frac{1}{\rho} \frac{\partial}{\partial \rho} \left( \rho \frac{\partial}{\partial \rho} \right) + (k^2 n^2 + h^2) - \frac{\nu^2}{\rho^2} \right] U = -j\omega\mu I\ell \delta(\rho - \rho_0)/\rho \quad (4)$$

Using the substitution

$$U = \frac{1}{\rho^{1/2}} V \quad (5)$$

Eq. (4) becomes

$$\frac{d^2 V}{d\rho^2} + q^2(\rho) V = -j\omega\mu I\ell \delta(\rho - \rho_0)/\rho^{1/2} \quad (6)$$

where

$$q^2(\rho) = k^2 n^2 - h^2 - \frac{\nu^2}{\rho^2} + \frac{1}{4\rho^2} \quad (7)$$

<sup>3</sup>Wentzell-Kramers-Brillouin.

Equation (6) is of the form to which the WKB method applies (Ref. 1). The WKB (approximate) solution is

$$V = \frac{1}{i 2 q^{1/2}(\rho_0) q^{1/2}(\rho) \rho_0^{1/2} \rho^{1/2}} \exp \left[ -i \int_{\rho_0}^{\rho} q(\rho) d\rho \right] \quad \text{for } \rho > \rho_0 \quad (8)$$

Substituting Eq. (5) and Eq. (7) into Eq. (3) yields

$$E = \frac{j\omega\mu I\ell}{(2\pi)^2} \int_{-\infty}^{\infty} dh \int_{-\infty}^{\infty} dv \frac{1}{i 2 q^{1/2}(\rho_0) q^{1/2}(\rho) \rho_0^{1/2} \rho^{1/2}} \exp \left[ -i \int_{\rho_0}^{\rho} q d\rho - jh(z - z_0) - jv(\phi - \phi_0) \right] \quad (9)$$

Letting  $h = ka$  and  $v = kb$ , Eq. (9) can be rewritten

$$E = j\omega\mu I\ell \frac{k}{(2\pi)^2} \int_{-\infty}^{\infty} da \int_{-\infty}^{\infty} db \frac{1}{j2 \left( n_0^2 - a^2 - \frac{b^2}{\rho_0^2} \right)^{1/4} \left( n^2 - a^2 - \frac{b^2}{\rho^2} \right)^{1/4} \rho_0^{1/2} \rho^{1/2}} e^{-jk\psi} \quad (10)$$

where

$$\psi = a(z - z_0) + b(\phi - \phi_0) + \int_{\rho_0}^{\rho} \left( n^2 - a^2 - \frac{b^2}{\rho^2} \right)^{1/2} d\rho \quad (11)$$

and since  $v$  is large ( $k$  is large), Eq. (7) becomes

$$q(\rho) = k \left( n^2 - a^2 - \frac{b^2}{\rho^2} \right)^{1/2} \quad (12)$$

Since  $k$  is large, Eq. (11) is of the form

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) e^{f(x, y)} dx dy \quad (13)$$

where  $f(x, y)$  is large, and saddle point integration may be used for evaluation (Ref. 2). The location of the saddle point is given by

$$\left. \begin{aligned} \frac{\partial \psi}{\partial a} = 0 &= z - z_0 + \frac{\partial}{\partial a} \int_{\rho_0}^{\rho} \left( n^2 - a^2 - \frac{b^2}{\rho^2} \right)^{1/2} d\rho \\ \frac{\partial \psi}{\partial b} = 0 &= \phi - \phi_0 + \frac{\partial}{\partial b} \int_{\rho_0}^{\rho} \left( n^2 - a^2 - \frac{b^2}{\rho^2} \right)^{1/2} d\rho \end{aligned} \right\} \quad (14)$$

Evaluation of Eq. (11) by saddle point integration, therefore, yields the final results

$$E = j\omega\mu I\ell \frac{1}{4\pi} \frac{1}{\rho_0^{1/2} \left( n_0^2 - a^2 - \frac{b^2}{\rho_0^2} \right)^{1/4} \rho^{1/2} \left( n^2 - a^2 - \frac{b^2}{\rho^2} \right)^{1/4}} \frac{1}{(\psi_{aa} \psi_{bb} - \psi_{ab}^2)^{1/2}} e^{-jk\psi} \quad (15)$$

where  $\psi$  is Eq. (11) and

$$\left. \begin{aligned} \psi_{aa} &= \frac{\partial^2 \psi}{\partial a^2} = \frac{\partial^2}{\partial a^2} \left[ \int_{\rho_0}^{\rho} \left( n^2 - a^2 - \frac{b^2}{\rho^2} \right)^{1/2} d\rho \right] \\ \psi_{bb} &= \frac{\partial^2 \psi}{\partial b^2} = \frac{\partial^2}{\partial b^2} \left[ \int_{\rho_0}^{\rho} \left( n^2 - a^2 - \frac{b^2}{\rho^2} \right)^{1/2} d\rho \right] \\ \psi_{ab} &= \frac{\partial^2 \psi}{\partial a \partial b} = \frac{\partial^2}{\partial a \partial b} \left[ \int_{\rho_0}^{\rho} \left( n^2 - a^2 - \frac{b^2}{\rho^2} \right)^{1/2} d\rho \right] \end{aligned} \right\} \quad (16)$$

Equations (15) and (16) agree with the amplitude and phase results of the geometrical optics representation derived in SPS 37-60, Vol. III. Since Eq. (1) is a scalar wave equation, it does not yield polarization information.

### 3. Conclusion

It has been shown that application of the WKB approximation to the scalar wave equation yields the same results obtained by geometrical optics. The scalar wave equation approach has one important advantage: the fields at points where the geometrical optics solution fail can now be evaluated. For instance, Eq. (15) fails at turning points, i.e., for

$$n^2 = a^2 + \frac{b^2}{p^2}$$

By going back to Eq. (10) and expanding the integrand, the fields at these turning points can indeed be obtained (Ref. 3). However, the scalar wave equation does not yield polarization information.

### References

1. Wait, J. R., "Electromagnetic Waves in Stratified Media," pp. 86-106, The MacMillan Co., New York, 1962.
2. Born, M., and Wolf, E., "Principles of Optics," pp. 750-751, Pergamon Press, Oxford, London, 1959.
3. Seckler, B. D., and Keller, J. B., "Asymptotic Theory of Diffraction in Inhomogeneous Media," *J. Acoust. Soc.* 31, pp. 206-216, February 1959.

## G. Spacecraft Antenna Research: Spacecraft Antenna Pointing Loss; Optimization Criterion for Illuminating Circular Antenna Apertures,

R. Holland

### 1. Introduction

For many applications the most desirable antenna pattern is a rectangle (Fig. 19), with power uniformly radiated into the cone  $\theta < \theta_0$ , but with no radiation outside

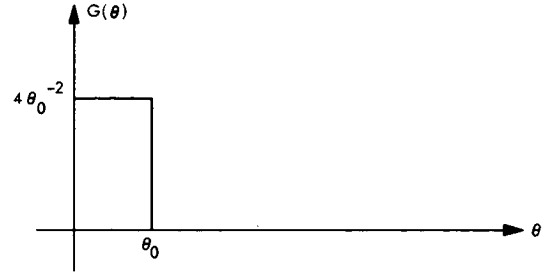


Fig. 19. Idealized antenna pattern of rectangular shape

that cone. This type of pattern is particularly useful when target location or antenna orientation is only known with a tolerance of  $\theta_0$ . Of course, this ideal pattern cannot be realized with a finite aperture. Generally speaking, however, how close one can come to this ideal depends only on  $ka\theta_0$ , where  $k$  is the radiation wave number and  $a$  is the antenna aperture radius.

The question of what aperture illumination is optimum with respect to approximating the ideal proves to be fairly difficult to answer. In this report, we shall describe a criterion which should be quite useful in comparing the effectiveness of different illuminations. We shall then illustrate by numerical examples the application of this criterion, and shall describe a class of illumination patterns which we believe to be fairly close to optimum.

### 2. Summary of Basic Formulas

Let us have a circular aperture of radius  $a$  in the  $z = 0$  plane. In this plane define  $r'$  as radius and  $\phi'$  as angle from the  $x$ -axis. Let the aperture be illuminated by an electric field which is polarized in the  $x$ -direction by a  $\phi'$ -independent excitation  $e'(r')$ . This field will be assumed to have a suppressed  $e^{-j\omega t}$  dependence. The electric far field is then given by

$$\mathbf{E}(r, \theta) = \frac{-\mathbf{i}_x j k e^{jkr}}{r} \int_0^a e'(r') J_0(kr'\theta) r' dr' \quad (1)$$

where  $r$  is the distance from the center of the aperture to the point of observation, and  $\theta$  is the angle by which the point of observation is off the  $z$ -axis. (We assume in this article that all significant radiation occurs at an angle close to the  $z$ -axis, so that we can always approximate  $\sin \theta$  by  $\theta$ .)

The power radiated by  $e'(r')$  into a unit solid angle at  $\theta$  is

$$P(\theta) = \frac{1}{2} Y_0 E_x(r, \theta) E_x^*(r, \theta) r^2 \quad (2)$$

where

$$Y_0 = \left( \frac{\epsilon_0}{\mu_0} \right)^{1/2} \quad (3)$$

Gain of an antenna at angle  $\theta$  is  $P(\theta)$  normalized by what  $P(\theta)$  would be if the antenna radiated isotropically:

$$G(\theta) = \frac{P(\theta)}{P_T/4\pi} \quad (4)$$

where  $P_T$  is the total power radiated. One can show by using Fourier transforms that  $P_T$  may be expressed as an integral of the far-field power flow or as an integral of the power flow through the aperture:

$$\begin{aligned} P_T &= \pi Y_0 \int_0^\infty E_x(r, \theta) E_x^*(r, \theta) r^2 d\theta \\ &= \pi Y_0 \int_0^a e'(r') e'^*(r') r' dr' \end{aligned} \quad (5)$$

If one combines Eqs. (1) to (5) it is possible to write  $G(\theta)$  as

$$\begin{aligned} G(\theta) &= \frac{2 \left( \int_0^a e'(r') J_0(kr'\theta) kr' dr' \right)^2}{\int_0^a e'(r') e'^*(r') r' dr'} \\ &= \frac{2 \left( \int_0^1 e(\beta) J_0(\beta ka\theta) \beta d\beta \right)^2 (ka)^2}{\int_0^1 e(\beta) e^*(\beta) \beta d\beta} \end{aligned} \quad (6)$$

by letting  $r' = \beta a$  and  $e'(\beta a) = e(\beta)$ . Thus,  $\beta$  becomes the normalized radial coordinate of the aperture:  $\beta = 1$  at  $r' = a$ .

### 3. Optimization Criterion

The basis of the optimization criterion we have chosen to employ is essentially a worst-case analysis. In particular, let us assume we are equally concerned about any signals in the cone described by  $\theta < \theta_0$ , and completely unconcerned about any signals outside that cone. Then an essential measure of the effectiveness of some antenna for this

application<sup>4</sup> is

$$G_{\min} = \min_{\theta \leq \theta_0} \{G(\theta)\} \quad (7)$$

An important consideration in discussing this problem is determining what happens should the aperture and its illumination be scaled down by a factor  $\alpha$ :

$$a \rightarrow a/\alpha$$

$$e(r') \rightarrow e(r'\alpha) \quad (8)$$

This operation causes  $G(\theta)$  to be altered as follows:

$$\begin{aligned} G(\theta) \rightarrow G_\alpha(\theta) &= \frac{2 \left( \int_0^{a/\alpha} e(r'\alpha) J_0(kr'\theta) kr' dr' \right)^2}{\int_0^{a/\alpha} e(r'\alpha) e^*(r'\alpha) r' dr'} \\ &= G(\theta/\alpha)/\alpha^2 \end{aligned} \quad (9)$$

Equation (9) shows that the pattern  $G(\theta)$  is broadened by  $\alpha$  (i.e.,  $\theta_0$  is increased by  $\alpha$ ), but the magnitude of  $G$  is decreased by  $\alpha^2$ . In other words, scaling down the aperture by  $\alpha$  expands the beamwidth by  $\alpha$ , but within the expanded beamwidth decreases  $G_{\min}$  by  $\alpha^2$ :

$$G_{\alpha, \min} = \min_{\theta \leq \alpha\theta_0} \{G_\alpha(\theta)\} = G_{\min}/\alpha^2 \quad (10)$$

A consequence of the preceding discussion is that the quantity  $G_{\min}\theta_0^2$  is invariant to operations of the type given by Eq. (8). Rather, this quantity depends only on the shape of  $e'(r'\alpha)$ , not on  $\alpha$  itself. Therefore, we can define a figure of merit which is dependent on the *shape* of the aperture illumination  $e'(r')$ , but not on the actual aperture size:

$$M\{e'(r')\} = \frac{1}{4} G_{\min}\theta_0^2 \quad (11)$$

The factor of 1/4 is introduced because it caused the ideal pattern of Fig. 19 to be characterized by  $M = 1$ . Equation (11) explicitly shows the interesting fact that

<sup>4</sup>In conventional terminology  $G_{\min}$ , expressed in dB, is simply antenna gain minus the pointing loss:

$$10 \log \{G_{\min}\} = 10 \log \{G(0)\} - 10 \log \left\{ \min_{\theta < \theta_0} G(\theta)/G(0) \right\}.$$

communication performance  $G_{\min}$  is fundamentally dependent on the inverse square of the pointing error  $\theta_0$ .

For finite apertures  $M$  will always be less than 1, provided  $de'(r')/dr'$  is nowhere infinite. In general, the optimum realizable  $e'(r')$  is that  $e'(r')$  which maximizes  $M$ . As one may infer from Eqs. (6) and (11), this optimum  $e'(r')$  will depend only on the product  $ka\theta_0$ :

$$e'_{\text{OPT}}(r') = e'\{ka\theta_0\} \quad (12)$$

A dependence between  $e'$  and  $ka\theta_0$  of the above type is mathematically described by saying that  $e'$  is a functional of  $ka\theta_0$ .<sup>5</sup> What we desire to do is to find  $e'\{ka\theta_0\}$  for all interesting values of  $ka\theta_0$ . This is not necessarily a simple task. However, once it has been accomplished, we can plot the relationship  $M_{\text{OPT}}(ka\theta_0)$ . Then, from Eq. (11), once any two of the three quantities ( $G_{\min}$ ,  $\theta_0$ ,  $ka$ ) are specified, we can at once determine the third quantity under optimum illumination and the associated optimizing field  $e'\{ka\theta_0\}$ . Typically, one might wish to fix  $G_{\min}$  and  $\theta_0$ , and then determine the minimum value of  $ka$  (i.e., aperture size) which is consistent with the required  $G_{\min}$  and  $\theta_0$ . This determination of the minimum satisfactory  $ka$  could be made by our procedure once  $M_{\text{OPT}}(ka\theta_0)$  and  $e'\{ka\theta_0\}$  have been found.

**a. Examples.** Before presenting any additional abstract theory, it seems advisable here to provide some specific examples relating to the material we have already covered. This will also give us an opportunity to introduce, in a systematic way, the relevant characteristics of some commonly used illumination patterns.

**Example 1—uniform illumination.** As a first illustration of the foregoing discussion, let us compute the figure of merit for the case of uniform illumination

$$\begin{aligned} e'(r') &= 1 & r' &\leq a \\ &= 0 & r' &> a \end{aligned} \quad (13)$$

If Eq. (13) is substituted into Eq. (6), one finds the gain for this case to be

$$G(\theta) = \frac{4J_1^2(ka\theta)}{\theta^2} \quad (14)$$

<sup>5</sup>Note that the optimum aperture illumination for a single value of  $ka\theta_0$  is the solution to the entire set of problems obtained by scaling the angle  $\theta_0$  by  $1/\alpha$  and the aperture size by  $\alpha$ .

Consequently, as a function of  $ka\theta_0$ ,  $M$  for uniform illumination is

$$\begin{aligned} M &= \frac{1}{4} G_{\min} \theta_0^2 \\ &= J_1^2(ka\theta_0) & ka\theta_0 &\leq j_{1,1} \\ &= 0 & ka\theta_0 &\geq j_{1,1} \end{aligned} \quad (15)$$

Note that for no values of ( $ka\theta_0$ ) is the uniform-illumination  $M$  higher than 0.34 ( $M$  reaches this peak at  $ka\theta_0 = 1.84$ ). Also note that  $M$  possesses the aforementioned invariance to shrinking the aperture  $ka$  by  $\alpha$  if one, at the same time, expands beamwidth of interest  $\theta_0$  by  $\alpha$ .

A characteristic of  $M$  which we again wish to emphasize is that it must be a function of the product  $ka\theta_0$ , but not of  $ka$  or  $\theta_0$  alone, for fixed illumination shape  $e(\beta)$ . This rule is an immediate consequence of Eq. (6) or alternatively of the invariance of  $M$  to the scale factor  $\alpha$ . One can observe in Eq. (15) the manifestation of this rule as it is reduced to the special case of uniform illumination.

**Example 2—rms optimized illumination.** Let us assume we have some desired far-field pattern  $E_{xd}(r, \theta)$ , such as that implied by Fig. 19, which we wish to simulate as closely as possible by illuminating an aperture of radius  $a$ . We can define an error power between the desired and actual far fields as

$$\begin{aligned} P_e &= \pi Y_0 \int_0^\infty (E_{xd}(r, \theta) - E_x(r, \theta)) \\ &\quad \times (E_{xd}^*(r, \theta) - E_x^*(r, \theta)) r^2 \theta d\theta \end{aligned} \quad (16)$$

in analogy with Eq. (5).

If the aperture radius  $a$  is not constrained to be finite, the theory of Fourier-Bessel transforms permits us to find the aperture illumination  $e'_d(r')$  which will give  $E_{xd}(r, \theta)$  exactly: Equation (1) may be written

$$E_{xd}(r, \theta) \left( \frac{-rk}{je^{jk r}} \right) = \int_0^\infty e''_d(R) J_0(R\theta) R dR \quad (17)$$

where

$$R = kr' \quad (18)$$

and

$$e''_d(R) = e''_d(kr') = e'_d(r') \quad (19)$$

The gain pattern of Fig. 19 is representative of the far-field pattern

$$\begin{aligned} E_{\theta d}(r, \theta) &= e^{jkr}/(kr) & \theta < \theta_0 \\ &= 0 & \theta \geq \theta_0 \end{aligned} \quad (20)$$

If Eq. (20) is substituted in Eq. (17) and the result multiplied by  $J_0(\rho\theta)\theta$  and integrated over  $0 \leq \theta < \infty$ , we obtain the aperture illumination associated with the ideal radiation pattern shown in Fig. 19:

$$e''_d(R) = \frac{j\theta_0 J_1(R\theta_0)}{R} \quad 0 \leq R = kr' < \infty \quad (21)$$

However, the actual aperture we must consider is of a finite radius  $a$ , and the realizable aperture illuminations thus are characterized by

$$e''(R) = 0 \quad R > ka \quad (22)$$

By virtue of Eq. (5), the error power in the far-field (Eq. 16) may be expressed as the aperture-plane integral

$$P_e = \pi Y_0 \int_0^\infty (e_d(r') - e(r'))(e_d^*(r') - e^*(r')) r' dr' \quad (23)$$

We can now see from Eqs. (21), (22), and (23) that the realizable aperture illumination which minimizes the error power is given by

$$\begin{aligned} e''(R) &= e''_d(R) = j\theta_0 J_1(R\theta_0)/R & R \leq ka \\ &= 0 & R > ka \end{aligned} \quad (24)$$

In other words, to minimize the far-field error power, one should compute the ideal aperture illumination for  $0 \leq r' < \infty$ , and then truncate this illumination at  $r' = a$ . (Note that as  $\theta_0 \rightarrow 0$  the shape of  $e''(R)$  in Eq. (24) approaches uniform illumination. This will, in fact, maximize performance in the absence of pointing error.)

The actual gain associated with Eq. (24) is obtained by substitution in Eq. (6):

$$G(\theta) = \frac{4I^2(ka\theta, ka\theta_0)(ka)^2}{(1 - J_0^2(ka\theta_0) - J_1^2(ka\theta_0))} \quad (25)$$

where

$$I(ka\theta, ka\theta_0) = \int_0^1 J_1(\beta ka\theta_0) J_0(\beta ka\theta) d\beta \quad (26)$$

As a function of  $ka\theta$ , the figure of merit  $M$  for this illumination becomes

$$M(ka\theta) = (ka\theta)^2 \min_{0 \leq \theta' \leq \theta_0} \left\{ \frac{I^2(ka\theta', ka\theta_0)}{1 - J_0^2(ka\theta_0) - J_1^2(ka\theta_0)} \right\} \quad (27)$$

In general,  $M(ka\theta)$  of Eq. (27) will reach a peak at some value of  $\theta < \theta_0$ . In other words, to come as close as possible to  $e'\{ka\theta\}$  for  $e'$  constrained to be of the form given by Eq. (24), one should take  $\theta_0$  a bit larger than  $\theta$ . This effect of  $M(ka\theta)$  reaching its maximum before  $\theta = \theta_0$  is a consequence of the fact that  $e''(ka)$  in Eq. (24) is rms-optimized, not figure-of-merit optimized: The two procedures are not equivalent, although we should hope one would be a good approximation to the other.

The rms-optimized illumination pattern of Eq. (24) has nulls at the Bessel function zeros

$$R\theta_0 = j_{1,i} \quad i = 1, 2, \dots \quad (28)$$

and extremal points at the Bessel function zeros

$$R\theta_0 = j_{2,k} \quad k = 1, 2, \dots \quad (29)$$

The far fields generated by Eq. (24) were evaluated numerically for aperture sizes selected so that  $ka\theta_0$  coincided with one of those nulls or external points:

$$ka\theta_0 = j_{1,1}, j_{2,1}, j_{1,2}, j_{2,2}, \dots \quad (30)$$

Numerical data were also obtained for

$$\begin{aligned} ka\theta_0 &= j'_{1,1} \\ &\frac{1}{2}(j'_{1,1} + j_{1,1}) \\ &\frac{1}{2}(j_{1,1} + j_{2,1}) \\ &\frac{1}{2}(j_{2,1} + j_{1,2}) \\ &\frac{1}{2}(j_{1,2} + j_{2,2}) \\ &\text{etc.} \end{aligned} \quad (31)$$

i.e., for  $ka\theta_0$  at  $J'_1(ka\theta_0) = 0$  and for  $ka\theta_0$  intermediate between the points given in Eq. (30).

Table 6 summarizes the conclusions reached from this numerical work. In this table,  $(ka\theta)_{\max}$  is the value of  $(ka\theta)$  at which  $M(ka\theta)$  is maximized. Also,  $G_{\min}$  is the minimal gain within the cone  $(ka\theta)' \leq (ka\theta)_{\max}$  and  $G(0)$  is the on-axis gain. The quantities  $G_{\min}$  and  $G(0)$  are normalized by  $(ka)^2$  in the table: Equation (14) indicates that  $G(0) = (ka)^2$  for uniform illumination. Consequently, the normalized figures,  $G_{\min}/(ka)^2$  and  $G(0)/(ka)^2$ , are measures of how much gain degradation is caused by deliberately spreading the beam over a cone instead of trying to focus it on the axis. This degradation may be seen to be very great if  $ka\theta_0$  is large.

It is interesting to observe that  $M(ka\theta)_{\max}$  is not a monotone increasing function of either  $(ka\theta_0)$  or  $(ka\theta)_{\max}$ . Especially at large  $(ka\theta_0)$ ,  $M(ka\theta)_{\max}$  is greatest when the illumination distribution is extremal at the aperture edge (i.e., when  $ka\theta_0 = j_{2,k}$ ).

Figures 20 and 21 illustrate the normalized gain  $G(\theta)/(ka)^2$  versus the normalized angle  $\theta/\theta_0$  for selected values of  $ka\theta_0$ .

*Example 3—Lambda-function illumination.* A frequently used generalization of the rms-optimized illumination

(Eq. 24) is the so-called lambda-function illumination

$$e''(R) = \frac{1}{2} j\theta_0^2 \Lambda_n([R^2\theta_0^2 + C]^{1/2}) \quad R \leq ka$$

$$= 0 \quad R > ka \quad (32)$$

where

$$\Lambda_n(u) = \frac{2^n n! J_n(u)}{u^n} \quad (33)$$

This choice of  $e''(R)$  reduces to the rms-optimized case if  $n = 1$  and  $C = 0$ . Note that if  $C$  is negative,  $\Lambda_n$  will involve a modified Bessel function over the portion of the aperture for which  $R\theta_0 < |C|^{1/2}$ .

If Eq. (32) is substituted into Eq. (6), we find the gain for lambda-function illumination is

$$G(\theta) = \frac{4 I^2(ka\theta, ka\theta_0) (ka)^2 (2n-1)}{2^{2(n-1)} n!^2 \left[ \frac{J_{n-1}^2(u)}{u^{2(n-1)}} + \frac{J_n^2(u)}{u^{2(n-1)}} \right]_{u=A}^{u=B}} \quad (34)$$

where

$$A = [(ka\theta_0)^2 + C]^{1/2}$$

$$B = C^{1/2}$$

Table 6. Characteristics of the rms-optimized far field for selected  $ka\theta_0$  values

$ka\theta_0$	$M(ka\theta)_{\max}$	$(ka\theta)_{\max}$	$G_{\min}/(ka)^2$	$G(0)/(ka)^2$
$j_{1,1} = 1.8411$	0.362	1.933	0.388 = 4.11 dB	0.982 = 0.08 dB
$\frac{1}{2}(j'_{1,1} + j_{1,1}) = 2.8364$	0.384	2.127	0.180 = 7.45 dB	0.891 = 0.50 dB
$j_{1,1} = 3.8317$	0.389	2.510	0.247 = 6.07 dB	0.639 = 1.95 dB
$\frac{1}{2}(j_{1,1} + j_{2,1}) = 4.4837$	0.401	3.139	0.163 = 7.88 dB	0.413 = 3.84 dB
$j_{2,1} = 5.1356$	0.461	3.852	0.124 = 9.06 dB	0.224 = 6.50 dB
$\frac{1}{2}(j_{2,1} + j_{1,2}) = 6.0756$	0.521	5.043	0.081 = 10.9 dB	0.082 = 10.9 dB
$j_{1,2} = 7.0156$	0.430	6.251	0.043 = 13.7 dB	0.043 = 13.7 dB
$\frac{1}{2}(j_{1,2} + j_{2,2}) = 7.7164$	0.480	6.752	0.042 = 13.8 dB	0.043 = 13.7 dB
$j_{2,2} = 8.4172$	0.617	6.818	0.053 = 12.8 dB	0.053 = 12.8 dB
$\frac{1}{2}(j_{2,2} + j_{1,3}) = 9.2953$	0.647	7.901	0.041 = 13.9 dB	0.053 = 12.8 dB
$j_{1,3} = 10.1735$	0.606	8.902	0.031 = 15.1 dB	0.067 = 11.7 dB
$\frac{1}{2}(j_{1,3} + j_{2,3}) = 10.8966$	0.650	9.371	0.029 = 15.4 dB	0.065 = 11.9 dB
$j_{2,3} = 11.6198$	0.702	9.935	0.028 = 15.5 dB	0.051 = 12.9 dB
$\frac{1}{2}(j_{2,3} + j_{1,4}) = 12.4717$	0.634	11.287	0.020 = 17.0 dB	0.034 = 14.7 dB
$j_{1,4} = 13.3237$	0.557	12.324	0.014 = 18.5 dB	0.014 = 18.5 dB
$\frac{1}{2}(j_{1,4} + j_{2,4}) = 14.0598$	0.619	12.836	0.015 = 18.2 dB	0.015 = 18.2 dB
$j_{2,4} = 14.7959$	0.746	13.124	0.017 = 17.7 dB	0.018 = 17.4 dB

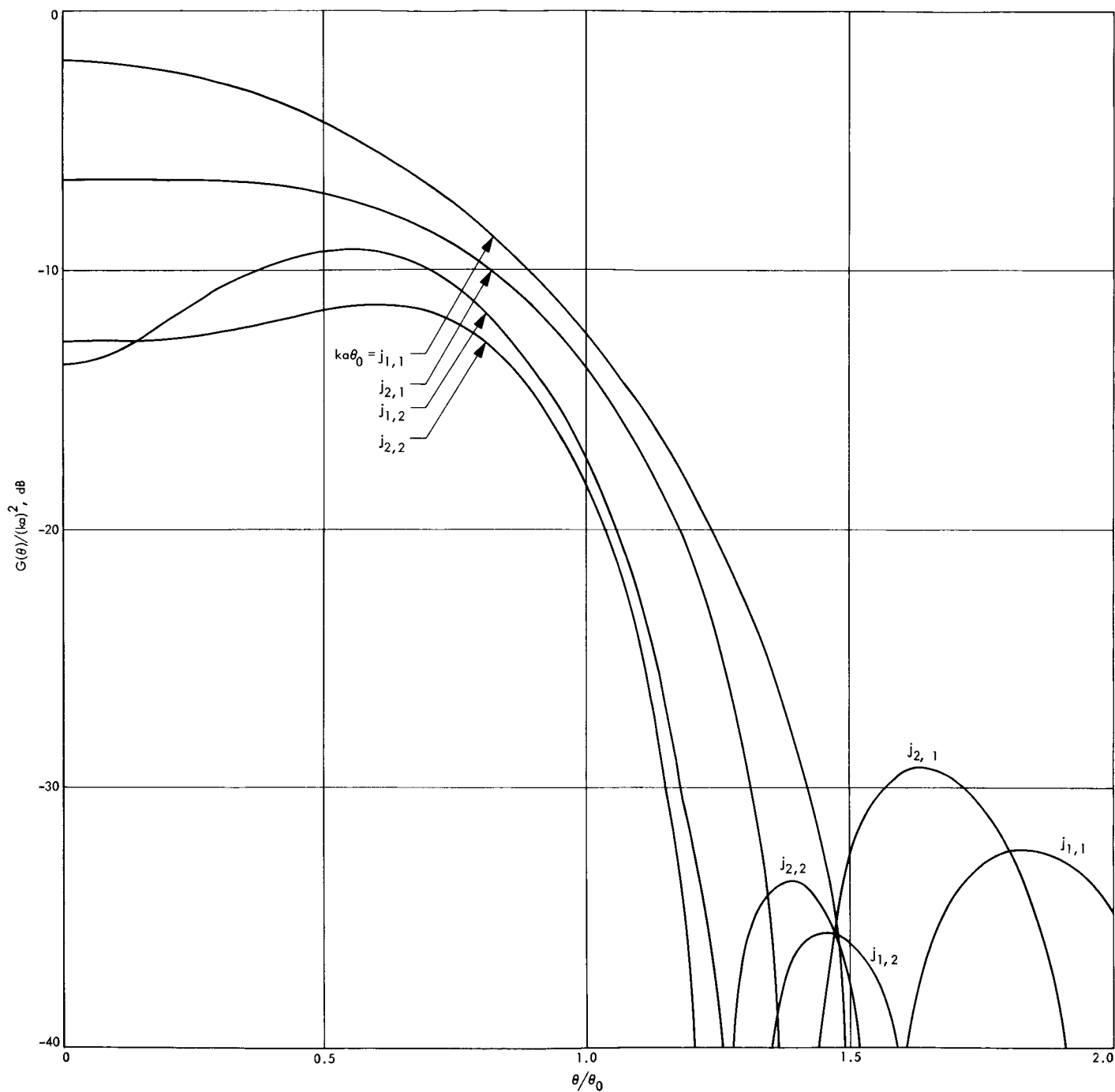


Fig. 20. Normalized gain  $G(\theta)/(ka)^2$  versus normalized angle  $\theta/\theta_0$  for rms-optimized illumination. The patterns for  $j_2k$  have flatter main beams but higher sidelobes than the patterns for  $j_{1,1}$ .

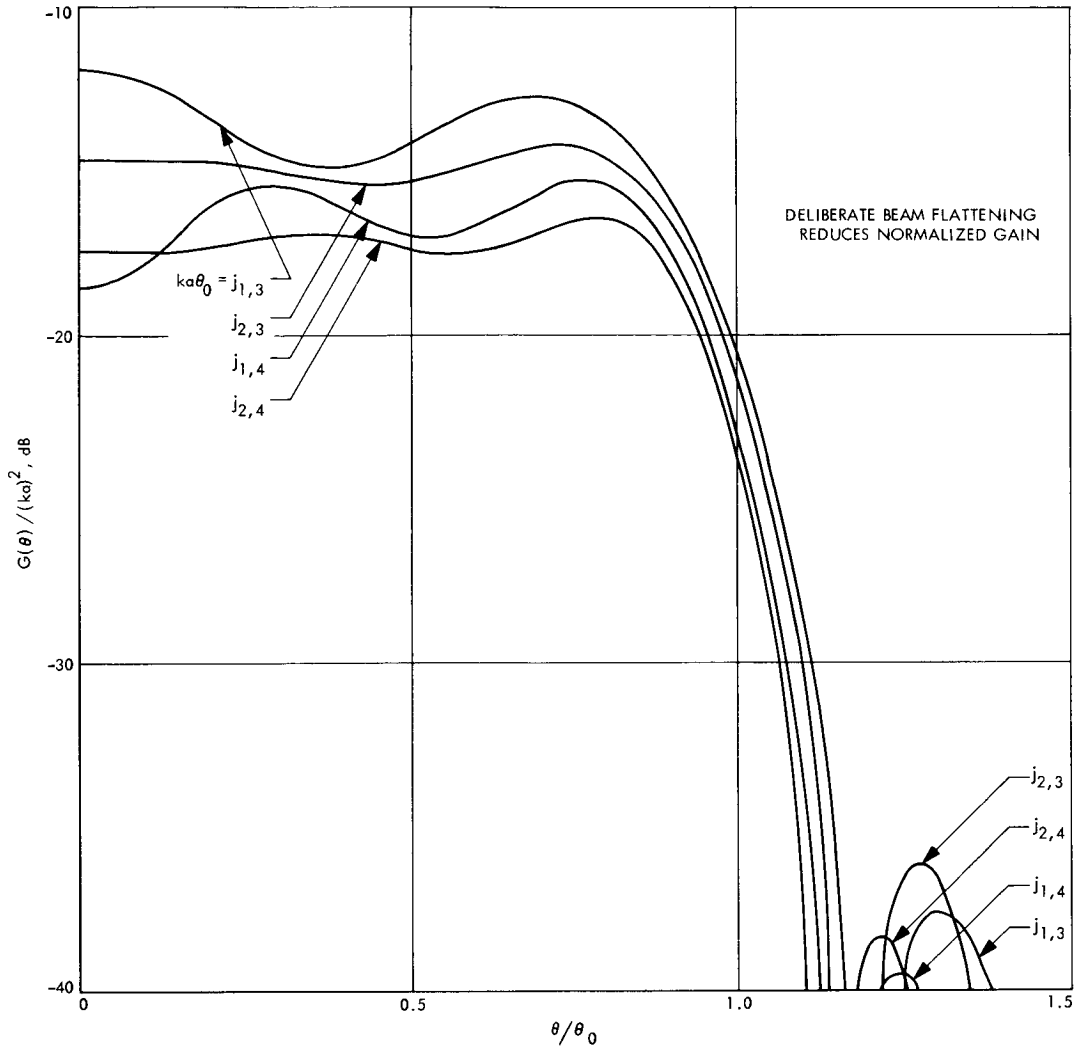


Fig. 21. Normalized gain versus angles for rms-optimized illumination. The apertures producing these patterns are larger than those producing the patterns of Fig. 20 for fixed  $\theta_0$

and

$$I(ka\theta, ka\theta_0) =$$

$$n! 2^{n-1} \int_0^1 \frac{J_n([\beta ka\theta_0]^2 + C)^{1/2}}{[(\beta ka\theta_0)^2 + C]^{n/2}} J_0(\beta ka\theta) (\beta ka\theta_0) d\beta \quad (36)$$

The figure of merit as a function of  $ka\theta$  for this illumination becomes

$$M(ka\theta) =$$

$$(ka\theta)^2 \min_{0 \leq \theta' \leq \theta} \left\{ \frac{I^2(ka\theta', ka\theta_0) (2n-1)}{2^{2(n-1)} n!^2 \left[ \frac{J_{n-1}^2(u)}{u^{2(n-1)}} + \frac{J_n^2(u)}{u^{2(n-1)}} \right]_{u=A}^{u=B}} \right\} \quad (37)$$

These lambda functions bear considerable resemblance to the Bessel functions from which they are derived, in that they are everywhere continuous, and often can be integrated in closed form. The lambda-function illumination given in this example, i.e., Eq. (32), has nulls at

$$R\theta_0 = (j_{n,i}^2 - C)^{1/2} \quad i = 1, 2, \dots \quad (38)$$

and extremal points at

$$R\theta_0 = (j_{n+1,k}^2 - C)^{1/2} \quad k = 1, 2, \dots \quad (39)$$

These formulas are very analogous to Eqs. (29) and (30) for the simple rms-optimized illumination.

Table 7 summarizes the properties of lambda illumination for some typical selections of the parameters  $n$  and  $C$ . In general, for a given  $ka\theta_0$ , lambda illumination leads to a figure of merit which is, at best, commensurate with that obtainable under rms illumination. In many cases, lambda illumination is quite inferior to rms illumination.

#### 4. Estimation of the Optimum Illumination and of $M_{OPT}(ka\theta)$

We shall now conclude this discussion by describing the specific numerical information we have obtained for estimating  $e'\{ka\theta\}$  and  $M_{OPT}(ka\theta)$ . First of all,  $M$  cannot exceed unity under any conditions, simply because of power conservation. However, if  $ka\theta$  is small,  $M$  is restricted by a yet more stringent limit. In particular, for axially symmetric antennas which do not employ supergain,  $G(\theta)$  can never exceed  $(ka)^2$ , and can only attain  $(ka)^2$  if  $\theta = 0$  and if the aperture has uniform illumination. Consequently, unless supergain techniques are applied,

$M$  as defined by Eq. (11) must be limited by

$$M(ka\theta) < \frac{1}{4}(ka\theta)^2 \quad (40)$$

If  $(ka\theta) < 2$ , Eq. (40) is a more severe restriction on  $M$  than its being bounded by unity. We refer to  $ka\theta \geq 2$  as the power-limited region and to  $ka\theta \leq 2$  as the supergain-limited or diameter-limited region. Figure 22 illustrates the respective behavior of the limits on  $M$  in those two regions.

In Fig. 22 we have also plotted discrete points for  $M(ka\theta)$  versus  $ka\theta$  as obtained numerically from uniform, rms-optimized, and lambda illumination. These discrete points appear to have an upper bound which is representable as a smooth function of  $ka\theta$ . This smooth bounding function is the best currently available approximation to  $M_{OPT}(ka\theta)$ . Furthermore, the corresponding estimate to  $e'\{ka\theta\}$  for each value of  $ka\theta$  is given by the illumination

Table 7. Characteristics of the far field resulting from selected lambda illuminations

$n$	$C$	$ka\theta_0$	$M(ka\theta)_{max}$	$(ka\theta)_{max}$	$G_{min}/(ka)^2$	$G(0)/(ka)^2$
1	1	$(j_{1,1}^2 - 1)^{1/2} = 3.6989$	0.389	2.497	0.250 = 6.02 dB	0.646 = 1.90 dB
1	1	$(j_{2,1}^2 - 1)^{1/2} = 5.0373$	0.478	3.904	0.125 = 9.03 dB	0.196 = 7.08 dB
1	1	$(j_{1,2}^2 - 1)^{1/2} = 6.9440$	0.306	6.819	0.026 = 15.8 dB	0.026 = 15.8 dB
1	1	$(j_{2,2}^2 - 1)^{1/2} = 8.3576$	0.567	7.522	0.040 = 13.9 dB	0.040 = 13.9 dB
1	0.5	$(j_{1,1}^2 - 0.5)^{1/2} = 3.7659$	0.389	2.542	0.241 = 6.18 dB	0.642 = 1.92 dB
1	0.5	$(j_{2,1}^2 - 0.5)^{1/2} = 5.0867$	0.469	3.840	0.126 = 9.00 dB	0.211 = 6.76 dB
1	0.5	$(j_{1,2}^2 - 0.5)^{1/2} = 6.9799$	0.373	6.533	0.034 = 14.7 dB	0.035 = 14.6 dB
1	0.5	$(j_{2,2}^2 - 0.5)^{1/2} = 8.3874$	0.603	7.180	0.047 = 13.3 dB	0.047 = 13.3 dB
1	-0.5	$(j_{1,1}^2 + 0.5)^{1/2} = 3.8964$	0.389	2.533	0.243 = 6.14 dB	0.637 = 1.96 dB
1	-0.5	$(j_{2,1}^2 + 0.5)^{1/2} = 5.1841$	0.455	3.758	0.129 = 8.89 dB	0.237 = 6.25 dB
1	-0.5	$(j_{1,2}^2 + 0.5)^{1/2} = 7.0511$	0.473	5.993	0.052 = 12.8 dB	0.052 = 12.8 dB
1	-0.5	$(j_{2,2}^2 + 0.5)^{1/2} = 8.4468$	0.607	6.546	0.052 = 12.8 dB	0.060 = 12.2 dB
1	-1.5	$(j_{1,1}^2 + 1.5)^{1/2} = 4.0227$	0.389	2.522	0.244 = 6.13 dB	0.632 = 1.99 dB
1	-1.5	$(j_{2,1}^2 + 1.5)^{1/2} = 5.2796$	0.443	3.696	0.130 = 8.86 dB	0.259 = 5.87 dB
1	-1.5	$(j_{1,2}^2 + 1.5)^{1/2} = 7.1217$	0.521	5.519	0.068 = 11.7 dB	0.071 = 11.5 dB
1	-1.5	$(j_{2,2}^2 + 1.5)^{1/2} = 8.5058$	0.570	6.260	0.054 = 12.7 dB	0.073 = 11.4 dB
0	0	$j_{0,1} = 2.4048$	0.390	2.405	0.270 = 5.69 dB	0.692 = 1.60 dB
0	0	$j_{1,1} = 3.8317$	0.000	0.000	0.000	0.000
0	0	$j_{0,2} = 5.5201$	0.024	1.325	0.053 = 12.8 dB	0.130 = 8.86 dB
0	0	$j_{1,2} = 7.0156$	0.000	0.000	0.000	0.000
2	0	$j_{2,1} = 5.1356$	0.388	2.604	0.227 = 6.44 dB	0.596 = 2.25 dB
2	0	$j_{3,1} = 6.3802$	0.410	3.413	0.140 = 8.54 dB	0.335 = 4.75 dB
2	0	$j_{2,2} = 8.4172$	0.466	4.629	0.087 = 10.6 dB	0.148 = 8.30 dB
2	0	$j_{3,2} = 9.7610$	0.458	5.486	0.060 = 12.2 dB	0.121 = 9.17 dB

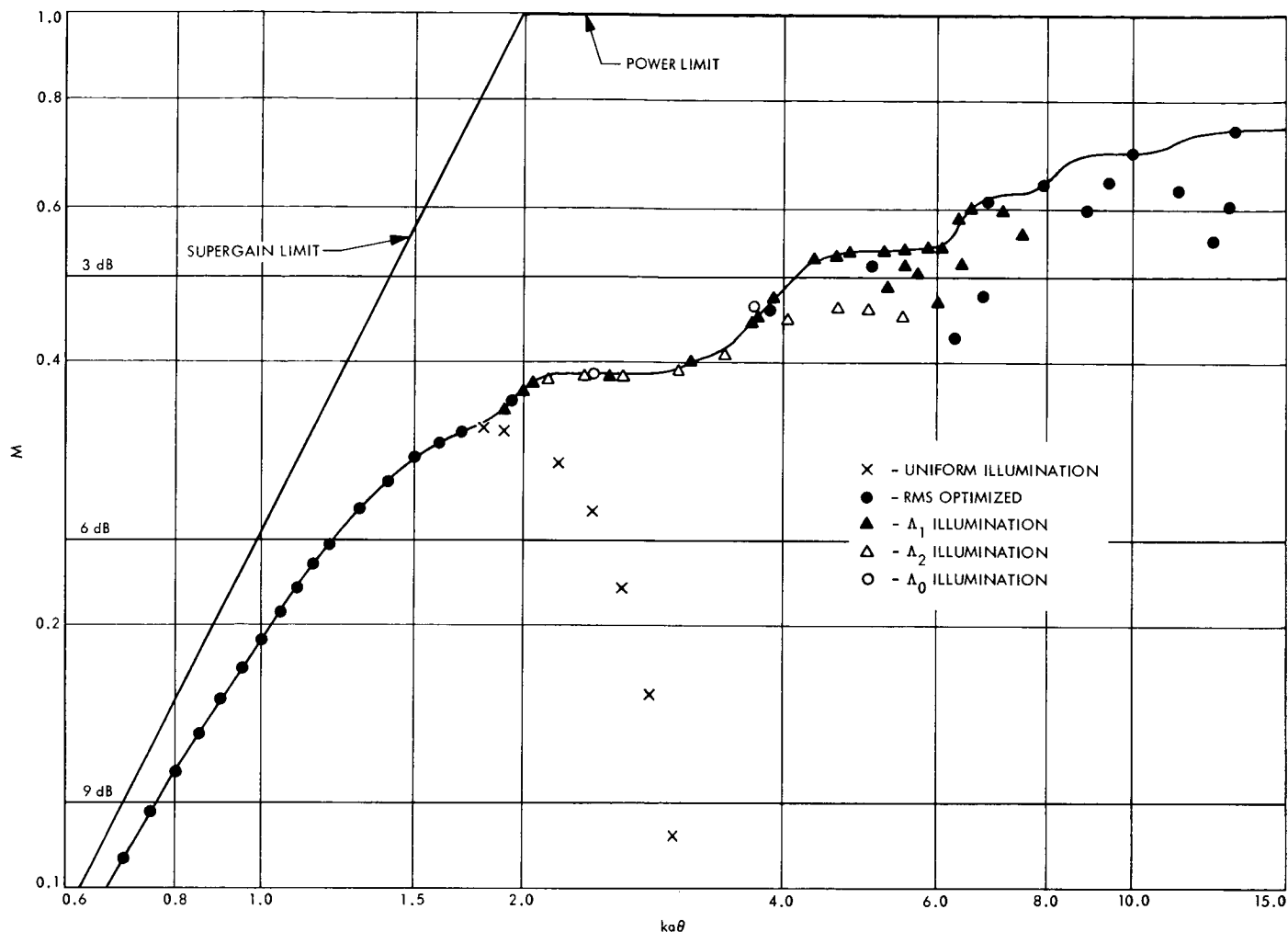


Fig. 22. Super-gain limit, power limit, and presently known bounding function for  $M$  as functions of  $ka\theta$

function which generated the discrete point closest to bounding function.

For  $ka\theta$  less than about 1.7, the rms-optimized and lambda illuminations are practically uniform over the aperture. Consequently, as one may see in Fig. 22, the figures of merit  $M$  are essentially identical for all the illumination shapes considered in this range of  $ka\theta$ . Also note that here the difference between these realizable  $M$ 's and the supergain limit on  $M$  is simply the conventionally defined pointing loss of the antenna.

A major unanswered question which Fig. 22 raises concerns the accessibility of the region of the  $M$  versus  $ka\theta$  plane between the bounding function and the supergain or power limit. This region is obviously quite wide for intermediate values of  $ka\theta$ , i.e.,  $1.5 < ka\theta < 6$ . It is our speculation that this region is largely forbidden to

realizable antennas. This statement does, however, merely represent a guess, and has not been mathematically supported.

## 5. Conclusion

For present systems with X-band frequencies and antennas 3 m in diameter, we have  $ka = 600$ . Current attitude controls are potentially capable of 0.002-rad accuracy. These figures correspond to a value of 1.2 for  $ka\theta$ . In this case Figure 22 shows that an attempt to shape the pattern with a sophisticated aperture illumination is not warranted. This is one immediately useful result. Figure 22 also shows, however, that any major increase in either antenna size or frequency or both, will require either beamshaping or better beam-pointing systems.

Additionally, we see that, although the rapid improvement due to increasing the antenna size drops off past

$ka\theta \sim 1.5$ , the value of  $M$  at this point is still about 5 dB below the fundamental limit, and 2 to 3 dB below the value achievable with really large antennas illuminated

by one of the solutions derived here. This fact will become important if an antenna concept is developed where large apertures or K-band systems are easily achievable.

## VI. Spacecraft Radio

### TELECOMMUNICATIONS DIVISION

#### A. TOPS Radio-Frequency Subsystem, A. W. Kermode

##### 1. Introduction

The purpose of the Thermoelectric Outer Planet Spacecraft (TOPS) radio-frequency subsystem (RFS) is to provide a design demonstration of a reliable one-way and two-way communications link between a spacecraft and earth throughout the lifetime of an outer planets mission. The functions that the RFS must perform with the TOPS spacecraft design are:

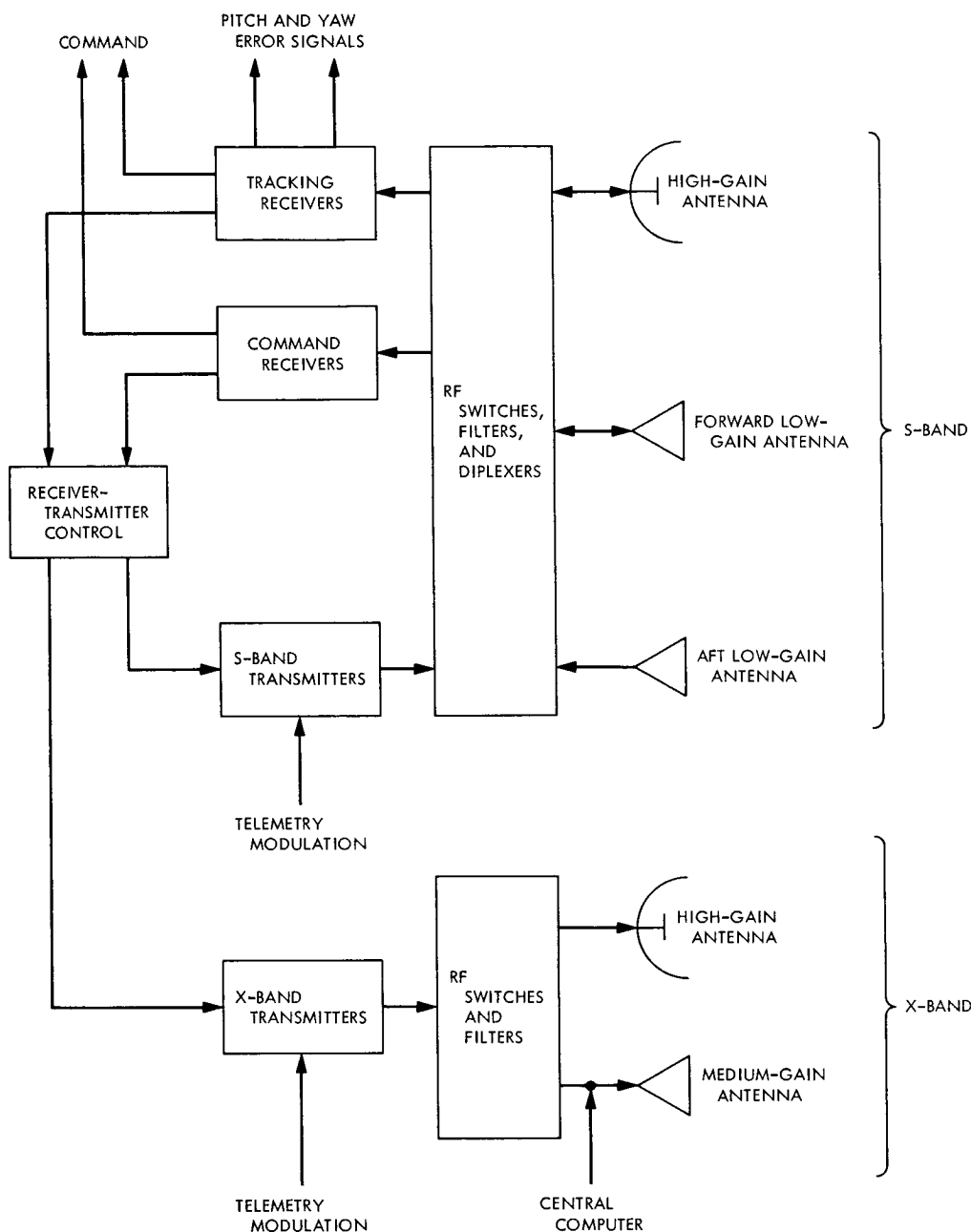
- (1) Receive the S-band signal transmitted to the spacecraft from the Deep Space Instrumentation Facility (DSIF).
- (2) Demodulate the received S-band signal for command and planetary ranging functions.
- (3) Coherently translate the frequency of the received S-band signal precisely by the turn around ratios of 240/221 for S-band operation, and 880/221 for X-band operation.
- (4) Develop pitch-and-yaw error signals from the S-band signal received via the high-gain monopulse antenna.
- (5) Transmit to the stations of the DSIF, a modulated S-band and X-band telemetry signal.

- (6) Modulate the transmitted signal with the detected ranging signal when operating in the planetary ranging mode.
- (7) Utilize on-board sensing for failure diagnostics, and appropriately switch in redundant system elements for failure correction as required.
- (8) Provide reliable performance throughout the mission lifetime ( $10^5$  h).

A Simplified functional block diagram of the RFS capable of implementing the previously listed functions is shown in Fig. 1. The RFS consists of the following redundant elements: command receivers, tracking receivers, S-band transmitters, and X-band transmitters. In addition it includes a receiver-transmitter control unit and inter-connecting hardware consisting of: filters, diplexer, RF switches, power-monitoring couplers, coaxial cables, and waveguide.

##### 2. Functional Description

During the first 250 days of the mission the S-band up-link signal will be received over the forward low-gain antenna and coupled to the operating command receiver through the diplexer-filter and RF switch. The command



**Fig. 1. TOPS radio-frequency subsystem**

signal, if present, will be demodulated and subsequently sent to the command subsystem. The S-band aft low-gain antenna will provide command reception capability in the rear hemisphere of the spacecraft in the event of spacecraft malfunction. After the first 250 days the S-band uplink will normally be over the high-gain antenna angle tracking receiver.

The S-band down-link cruise telemetry signal will be via the low-gain antenna for approximately the first 100

days of the mission. From approximately 100 to 250 days the cruise telemetry link will be via the X-band transmitter and the X-band single-axis gimbaled medium-gain antenna. The X-band medium-gain antenna will also be used for downlink telemetry during all midcourse maneuvers. After 250 days, the cruise telemetry link will be via the high-gain antenna at S-band. In addition to the S-band monopulse uplink and S-band downlink, the high-gain antenna feed is capable of transmitting at X-band. The X-band transmitter and X-band high-gain antenna

will be used to provide the periodic high data rate dump and for transmission of real-time imaging data. All high-gain antenna functions will utilize a common 14-ft reflector, with a composite multifrequency cassegrain feed system.

The RF tracking receiver will be coupled to the S-band high-gain antenna via a monopulse feed network. Pitch-and-yaw error signals will be developed in the tracking receiver, during two-way communications operation, for use by the attitude-control subsystem in maintaining the narrow-beam high-gain antenna earth-pointed. The objective of the RF pointing system is to maintain the antenna pointing losses to less than 0.3 dB at X-band (0.09 deg for a 14-ft parabola antenna).

The voltage-controlled oscillator output from the operating receiver will be coupled to the transmitter exciters through the transmitter control unit for use during coherent two-way operation. An auxiliary oscillator will provide the frequency source for one-way noncoherent operation.

The S-band and X-band transmitters will be capable of simultaneous or independent operation and will be coupled to the appropriate antennas by RF switches as required throughout the mission.

### 3. New RFS Requirements

The TOPS RFS requirements that are new or different from the previous *Mariner* spacecraft RFS requirements are:

- (1) The use of low-noise temperature command receivers.
- (2) The use of RF tracking receivers for minimization of antenna pointing losses.
- (3) The use of an X-band downlink, with simultaneous S-band and X-band operational capability.
- (4) Long mission lifetime ( $10^5$  h), which requires the use of block redundant receivers as well as the previously used block redundant exciters and traveling wave tube (TWT) power amplifiers.

The first three items listed above are required to provide overall communication link performance. The latter results in new problems, one of which is TWT cathode wearout occurring before end of the mission.

The *Mariner* Mars 1969 RFS estimated parts failure rate data was utilized to ascertain the impact of current

hardware characteristics on the TOPS RFS reliability. The results of this reliability study are summarized in Table 1. The *Mariner* Mars 1969 estimated failure rate data can be considered as pessimistic because of the confidence level to which the parts failure rate estimates are known. However, it does serve as a guide to show that an overall improvement in reliability is required for TOPS. Reliability improvement can be achieved through: (1) use of piece parts that have lower failure rates and more thorough screening; (2) use of fewer piece parts within the RFS modules; (3) operational use of parts at lower stress ratios; and (4) use of functional redundancy.

### 4. Receivers

The reception of emergency uplink commands at any time during the outer planet missions (out to 30 AU) requires the use of a low-noise command receiver coupled to a low-gain antenna providing hemispherical coverage, with use of a 210-ft ground station antenna transmitting 400 kW of RF power. A receiver noise temperature of less than 300°K is required to provide this function. The receiver will consist of a reliable low-noise amplifier coupled with a reliable S-band receiver of advanced design.

The monopulse-tracking receiver will be of a similar and compatible advanced design, with a receiver noise temperature of less than 630°K. The additional modules required to provide the antenna pointing capability will be added to the S-band command receiver design, thus also providing command and ranging capability.

### 5. Transmitters

The S-band transponder will provide the S-band and X-band exciters with a signal coherently related to the

**Table 1. RFS estimated failure rate and reliability data**

Component	Failure rate	Probability of success		
		<i>Mariner</i> Mars 1969 (5100-h mission), 1 unit	TOPS ( $10^6$ -h mission)	
			1 unit	2 units
Receiver and power supply	1.15	0.94	0.31	0.66
	0.58 <sup>a</sup>	—	0.56 <sup>a</sup>	0.88 <sup>a</sup>
Exciter and power supply	0.76	0.96	0.46	0.81
	0.38 <sup>a</sup>	—	0.68 <sup>a</sup>	0.94 <sup>a</sup>

<sup>a</sup>Based on improved failure rate data, i.e., 0.5 *Mariner* Mars 1969 failure rate.

S-band uplink. In the absence of an uplink signal the receiver-transmitter control unit will generate the RF signal. The X-band transmitter will provide the capability of transmitting data at rates up to ten times the S-band rate, depending on ground station weather conditions. Simultaneous coherent S-band and X-band operational capability are desired for the charged particle and radio science experiments, as well as for increased navigation accuracy (goal of 1-m accuracy). Each transmitter system (S-band and X-band) will provide partial functional redundancy for the other in the event of a catastrophic failure in one of the downlink systems.

Transmitter reliability studies were performed to determine the effects of cathode wearout on overall reliability for the TOPS mission and to compare the use of redundant single-cathode TWTs with multiple-cathode TWTs. Projected estimates of failure rates for RF hardware of the 1975 time period (assuming aggressive advanced development) were included in all calculations. The following items were included in the studies: (1) the mean and standard deviation of TWT cathode wearout lifetime; (2) the mission lifetime; and (3) the effects on overall transmitter reliability as a result of connecting (switching) multiple units to a common antenna. The results of the single-cathode TWT reliability study are summarized in Table 2. The tentative conclusion of the multiple-cathode TWT study is that the reliability resulting for a given number of redundant, single-cathode TWTs can generally be achieved with one fewer multiple-cathode TWTs.

Several design approaches are being pursued to resolve the TWT cathode wearout problem. The first is the use of improved longer life cathodes. Another approach is the use of a TWT that has the capability to switch a new cathode into position upon depletion or failure of the operating element. A third approach, at S-band, is the

**Table 2. TWT and power supply estimated reliability**

No. of units	Mean cathode wearout life, h <sup>a</sup>	Estimated reliability <sup>b</sup> (10 <sup>5</sup> -h mission)
3	40,000	0.50
3	45,000	0.62
3	50,000	0.72
4	40,000	0.83
4	45,000	0.89
4	50,000	0.92
5	45,000	0.97
<sup>a</sup> Standard deviation = 5000 h.		
<sup>b</sup> TWT/power supply failure rate = $10 \times 10^{-6}$ failures per hour.		

**Table 3. Hybrid S-band transmitter (TWT and solid state) estimated reliability**

No. of units		Estimated reliability	
TWT	Solid state	Solid-state transmitter estimated failure rate	
		$5 \times 10^{-6}$ failures per hour	$10 \times 10^{-6}$ failures per hour
3	2	0.99	0.98
2	2	0.98	0.95
TWT cathode mean wearout life = 45,000 h			
Standard deviation = 5,000 h			
TWT/power supply failure rate = $10 \times 10^{-6}$ failures per hour			

use of an all solid-state transmitter or a hybrid combination of solid-state and TWTs. Current S-band solid-state technology does not provide the overall efficiency achievable with TWTs. However, in the 1975 time period high-power S-band solid-state devices are expected to be more competitive with the TWT. The results of a hybrid transmitter reliability study are summarized in Table 3. The tentative conclusion of the study is that the use of solid-state transmitters provides higher reliability over that obtained with single-cathode TWTs. The solid-state transmitter not only eliminates the cathode wearout phenomenon but also requires a power supply of less complexity than the high-voltage TWT supply.

Cathode wearout is not expected to be a problem for the X-band TWT because of the low-duty usage throughout the TOPS mission. However, repeated on-off cycling, coupled with the dormancy factor in a space environment, may affect overall reliability and will require investigation.

A study was performed to show the expected reliability for the dual redundant X-band TWTs, for a range of failure rates. The results of the study are summarized in Table 4. The tentative conclusion of the study is that dual

**Table 4. Estimated X-band TWT/power supply reliability (2 units)**

TWT/PS estimated failure rate (10 <sup>-6</sup> failures per hour)	Probability of success (10 <sup>5</sup> -h mission)
10	0.98
15	0.96
20	0.94
TWT cathode mean wearout life 45,000 h	
Standard deviation 5,000 h	
Total estimated operating time 17,000 h	

redundant X-band TWTs will be acceptable for the TOPS mission provided that the lower failure rates ( $10$  to  $15 \times 10^{-6}$  failures per hour) are achieved, and no long-term space environment storage problems are encountered.

The power output level of the S-band TWT is currently expected to be a nominal 20 W. TWT technology is expected to provide a gain of greater than 30 dB with an overall efficiency greater than 43%.

The X-band TWT, as presently conceived, will provide a dual RF output power level. The RF power levels have not been established firmly as yet. However, levels of 20 and 40 W are under consideration, so that maximum transmitter power, commensurate with available spacecraft raw power at various phases of the mission, can be utilized to enhance performance of the high data rate

downlink. A gain greater than 30 dB with an overall efficiency greater than 35% is expected.

## 6. Summary

Requirements for the outer planet missions exceed present RFS capabilities in terms of lifetime, RF performance, dc power requirements, volume and weight. A comparison of the *Mariner* and TOPS RFS characteristics are summarized in Table 5. The significant differences in the TOPS RFS over the previous *Mariner* type are: (1) The receiver noise temperature will be lowered by greater than 5.5 dB; (2) the transmitter power at X-band will be increased by at least 3 dB over the previously used S-band level; (3) the maximum raw dc power requirement will increase by a factor of 1.35; and (4) the weight increase will be a factor of 1.7, with a volume increase of less than 1.5, while the RFS complexity increases by a factor of 3.5.

**Table 5. Comparison of *Mariner* and TOPS RFS characteristics**

Parameter	<i>Mariner</i> Mars 1964	<i>Mariner</i> Venus 1967	<i>Mariner</i> Mars 1969	<i>Mariner</i> Mars 1971	TOPS
Mission lifetime, h	5600	3200	5100	>6800	100,000
Nominal receiver noise temperature (Input to RFS case), °K	2700	2700	1670	1670	<300
Nominal transmitter power level, CW W S-band X-band	5/10 —	5/10 —	10/20 —	10/20 —	20/—/20/— —/20/20/40
Raw power requirements, W dc ac	38.8/57.8 20.2	33.0/61.8 20.5	59.7/99.0 32.3	59.7/99.0 32.3	65/78/126/134 <sup>a</sup> —
Weight, lb	34.3	32.9	53.7	53.7	91.7
Volume, in. <sup>3</sup>	1750	1750	1340	1340	1980
<sup>a</sup> Assumes all dc distribution system.					

## VII. Spacecraft Telecommunications Systems

### TELECOMMUNICATIONS DIVISION

#### A. Digital Command System Development,

C. R. Tegnalia

##### 1. Introduction

An all-digital, single-channel command subsystem has been suggested for upcoming missions<sup>1</sup> (SPS 37-59, Vol. III, pp. 68-70). One realization of such a system is presently being evaluated by computer simulation and breadboard testing. Some preliminary experimental data are available and further tests are underway.

##### 2. System Description

An algorithm for an all-digital command subsystem was described in SPS 37-59, Vol. III. Figure 1 shows an efficient implementation of this algorithm that was chosen for breadboard testing. The system consists basically of a digital correlator that simultaneously correlates the incoming subcarrier signal against 16 phase-offset estimates and stores the results in a long shift register R1. The number of subcarrier phase estimates is determined by the sampling frequency (see *Chapter IV-A*), which in turn is determined by the desired phase resolution. For the

present preliminary version, the indicated resolution was selected.

The contents of registers R2 (the first word of register R1) and R3 are compared in order that the largest correlation value may be gated into R3 for storage. The location in R1 of this largest value determines the best subcarrier phase estimate, and after a 5-bit accumulation time (see *Chapter IV-A*), the C3 counter is set to the proper phase.

The basic system hardware performs a similar operation to acquire bit sync, for which a change is made in the correlation signal. The proper correlation signal for each of the two acquisitions is provided by a function generator that is controlled by the mode-control sequencer. By operating in different correlation modes, the basic system hardware performs all the functions of synchronization and detection.

Once the system achieves synchronism, the basic hardware is used to detect data. Logic decodes the counter chain to locate the in-phase correlation value in R1, and its sign is output as data after an accumulation time of one command bit. Bit sync is provided directly from counter C4. Signal monitor logic forms the two's complement, when necessary, to remove the data modulation from the correlation values so that a sum of absolute values is

<sup>1</sup>Private communication with R. M. Goldstein and E. C. Posner of the JPL Communications Systems Research Section and L. A. Couvillon of the JPL Spacecraft Telemetry and Command Section.



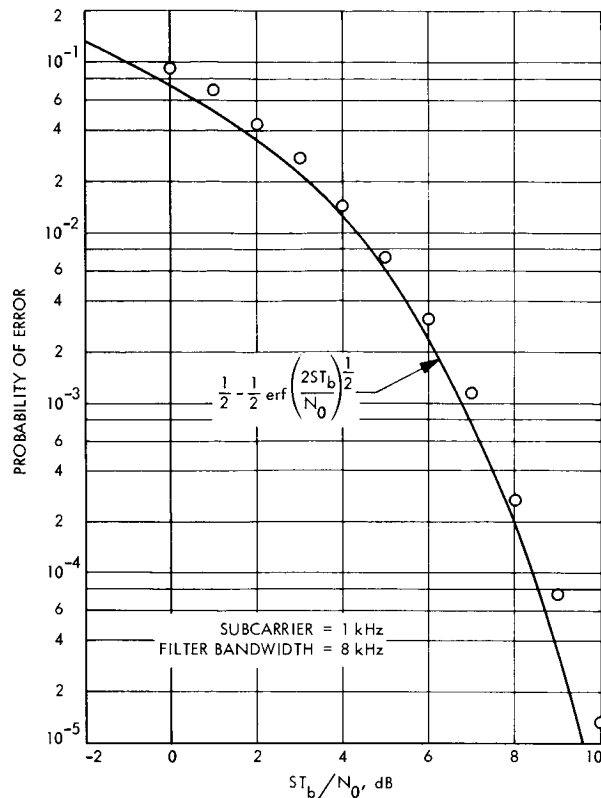


Fig. 2. Experimental bit error rates

formed. This statistic is accumulated to test for the presence of signal.

### 3. System Development

A breadboard model of the system shown in Fig. 1 is currently in operation and performs successfully at data rates up to 64 bits/s. The model was built entirely from Series 74 integrated logic circuits except for the analog-to-digital converter (A/DC) and metal-oxide semiconductor shift registers, which impose a theoretical upper limit of 80 bits/s. Limited testing of the breadboard model under noisy signal conditions has just begun; however, some preliminary bit error-rate data have been taken and are shown in Fig. 2. The data show a combined quantization and filtering loss of about 0.5 dB of signal with respect to the theoretical performance at  $ST_b/N_0 = 10$  dB. Perfect frequency synchronization was provided.

Some computer simulation has been performed in an effort to predict or improve system performance. Since the system is digital, precise simulation of performance is obtained by processing a digitized signal, equivalent to that at the output of the system A/DC, by digital computer. A computer simulation of subcarrier acquisition,

performed by N. A. Burow of the JPL Spacecraft Telecommunications Systems Section, produced the results shown in Fig. 3. The histogram shows the probability of acquiring subcarrier with a given phase error. Since the present system correlates against 16 offset estimates of subcarrier phase, errors are restricted to multiples of  $\pi/8$ . Another simulation performed by J. Springett also of the JPL Spacecraft Telecommunications Systems Section demonstrated that a signal-to-noise ratio loss of less than 0.4 dB results at threshold from using a 4-bit rather than 12-bit A/DC.

### 4. Further Work

Additional experimental tests are either planned or are underway. Acquisition probabilities for both subcarrier and bit sync are being evaluated, as is the behavior of the signal-monitor device. Computer simulation will be performed to determine the effect of analog-to-digital conversion on threshold values. In addition, a tracking feature is to be added to the system whereby initial phase errors or doppler shifts may be tracked out. In the more distant future, a parallel-registers version of the system may be employed to allow for much higher subcarrier rates.

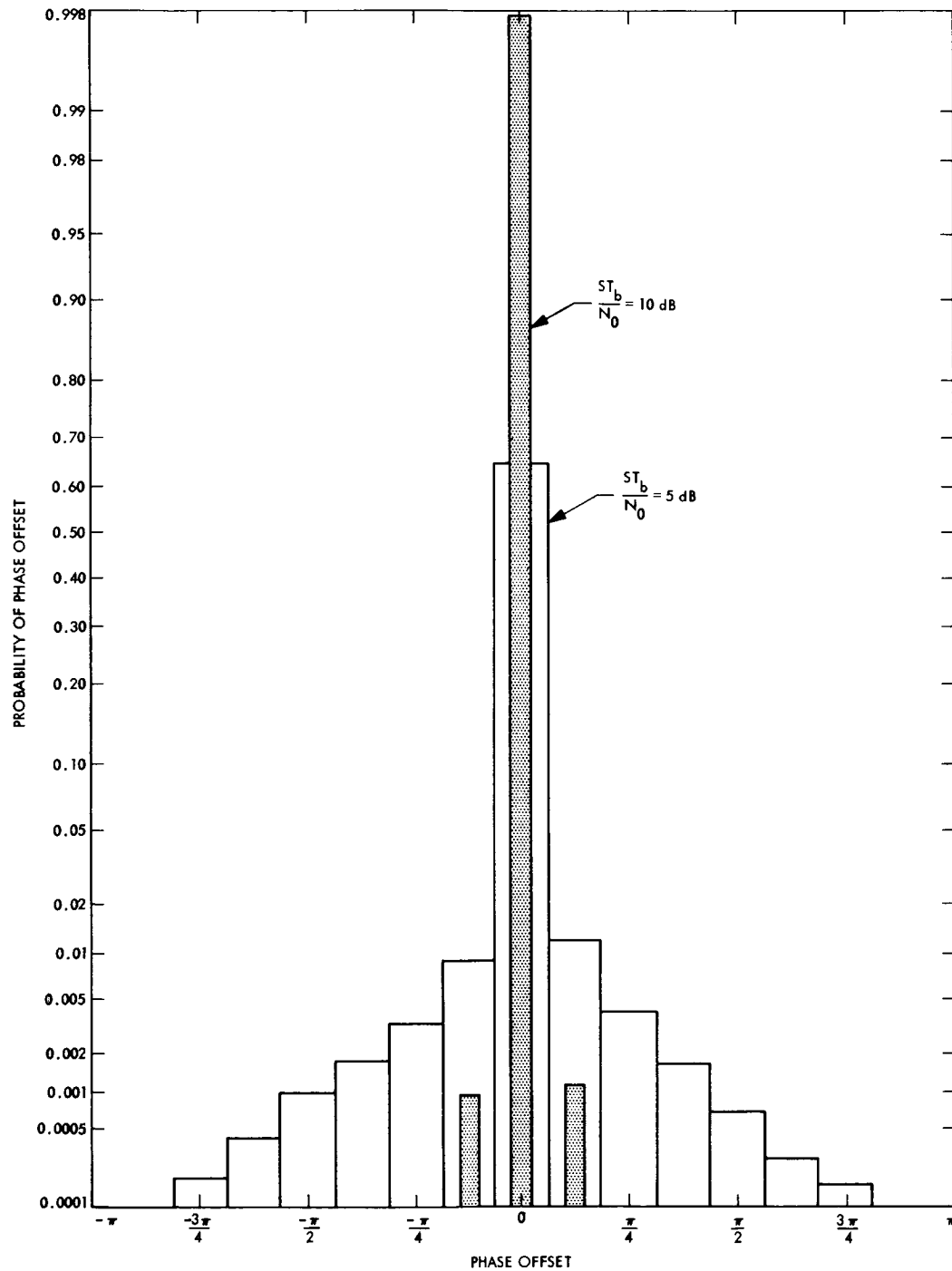


Fig. 3. Simulation of subcarrier acquisition

## VIII. Spacecraft Power

### GUIDANCE AND CONTROL DIVISION

#### A. A 400-W dc-to-ac Inverter for TOPS Power

Subsystem, D. Hopfer, C. McLyman, and J. Loughin<sup>1</sup>

##### 1. Introduction

An inverter has been designed to support an ac power distribution study in the Thermoelectric Outer-Planet Spacecraft (TOPS) Power Subsystem Development Program. It will receive pre-regulated 30-Vdc  $\pm 1\%$  power from a radioisotope thermoelectric generator (RTG) and deliver up to 400 W of ac power for distribution to the various spacecraft loads. The desired output of the inverter is a 50-V rms square wave with rise and fall times

between 1 and 5  $\mu$ s and an efficiency of 92%. The rise and fall times of the output are controlled to keep the electromagnetic interference generated on the distribution lines down to a tolerable level.

##### 2. Design Description

As can be seen from the block diagram (Fig. 1), the circuitry provides for input and a frequency synchronizing output overload protection. Under normal operation, the oscillator is driven by a synchronous signal at 4096 Hz but it is also capable of free running at approximately 3700 Hz.

The oscillator (Fig. 2) is a modified Jensen inverter with an additional transistor (Q3) for starting; C4, R4, and

<sup>1</sup>Space Systems Division, General Electric Co., Philadelphia, Pa.

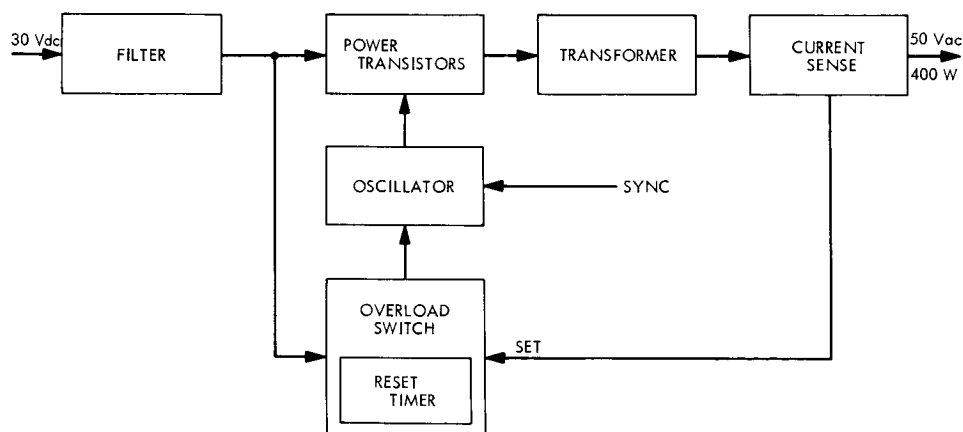


Fig. 1. Inverter block diagram

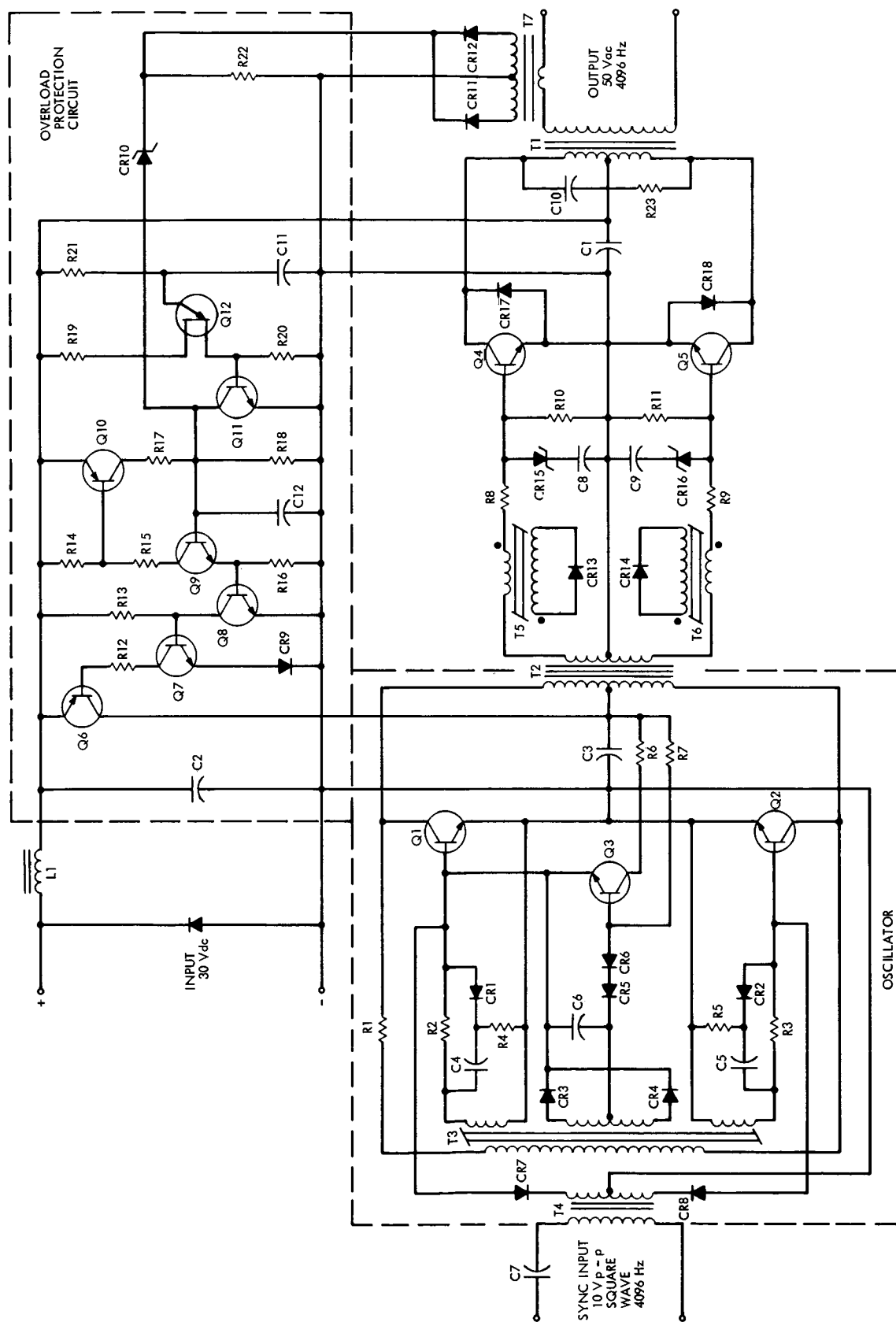


Fig. 2. Schematic diagram of 400-W dc-to-ac inverter

CR1 are added to improve the switching characteristics of the oscillator. The output of the oscillator provides base drive for the power switching transistors Q4 and Q5.

T5 is a volt-second delay transformer which prevents Q4 from turning on before Q5 turns off. The delay time is set equal to the turn-on time minus the turn-off time of the power transistors by adjusting the number of turns. In this design, the required delay time was  $1.6 \mu\text{s}$ . The magnetizing current required by T5 during the delay time is conducted through R10 bypassing the power transistor. During the turn-off phase, T5 presents a negligible impedance in the drive circuit due to conduction through CR13, which also provides the voltage required to reset T5.<sup>2</sup> T6 is also a volt-second delay transformer that performs the same function for power transistor Q5 as the T5 volt-second device does for transistor Q4.

For overload protection, the load current is sensed by the current transformer T7. When an overload occurs, transistor Q6 will turn off, removing power from the oscillator and thus turning off the inverter. Due to the latching nature of Q9 and Q10, the inverter will remain off until a reset pulse is received from the unijunction timing circuit (Q12). The inverter will continue to pulse on and turn off until the overload is removed.

### 3. Test Results

The inverter has been breadboarded and temperature tested from  $-20$  to  $+80^\circ\text{C}$  with satisfactory results (Table 1). The switching time of the power transistors plus the effect of leakage inductance in the power transformer was sufficient to give the desired rise and fall times on the output voltage without any additional waveshaping.

<sup>2</sup>Development of this circuitry has been reported in SPS 37-60, Vol. III, pp. 66-69.

Table 1. Inverter test results<sup>a</sup>

Temperature, °C	P <sub>out</sub> , W	V <sub>out</sub> , V rms	Efficiency, %
-20	200	51.1	92.2
-20	300	50.5	93.4
-20	400	49.7	93.2
+20	200	51.0	92.0
+20	300	50.3	92.8
+20	400	49.6	92.9
+80	200	50.8	91.7
+80	300	50.1	92.1
+80	400	49.1	92.2

<sup>a</sup>V<sub>in</sub> = 30 Vdc.

The inverter was tested for overload protection by applying both a slight overload and short circuit to the output. In both cases, the inverter cut off in less than  $5 \mu\text{s}$  after application of overload.

Table 1 gives a summary of the test data. No significant change was observed in output voltage or efficiency when the inverter was operated in the normal synchronized mode and the free running mode.

## B. A Test Program for the Investigation of Gravity Effects on Silver-Zinc Cells, G. L. Juvinal

### 1. Introduction

The effects of low-gravity environments on the performance of electrochemical cells have not yet been defined. Although anomalies that may be due to the low-gravity environment have been observed in the behavior of flight batteries, more data are required to resolve the problem. In support of a proposed flight experiment, a breadboard low-gravity battery test unit has been fabricated and is undergoing tests at JPL.

The capabilities of the breadboard unit include measurement of the limiting current of smooth flooded zinc anodes, measurement of the limiting current and capacity of conventional silver-zinc battery cells, and a photographic study of the formation and behavior of bubbles at the surface of silver and zinc electrodes during flight. The photographic results are correlated with electrical measurements of capacity and limiting current.

The measurement of the limiting current and capacity of conventional silver-zinc cells is described here. The other functions of the breadboard unit have been described previously (SPS 37-52, Vol. III, pp. 38-44; SPS 37-55, Vol. III, pp. 107-109). Earlier work at JPL has also investigated the effects of sustained high  $g$  levels in the range of 1 to 75  $g$  (SPS 37-55, Vol. III, pp. 107-109).

The purpose of the limiting current (polarization) and capacity measurements is to investigate the degradation of conventional silver-zinc cells in low-gravity environments. For this reason, the cells are subjected to charge-discharge cycling in conjunction with the actual performance measurements.

### 2. Experimental Approach

Yardney Electric Corp. HR5-DC7A cells were used for the tests. The cells are conventional 5A-h silver-zinc cells,

containing six positive and seven negative plates. The positive plates are wrapped with four layers of C-19 cellophane.

The tests involve an initial capacity and polarization measurement, followed by a series of five charge-discharge cycles, and a final capacity and polarization measurement. Cell degradation can be assessed by comparing the initial and final capacity and polarization measurements. Provisions are made for sequentially testing six cells with a calibrating step between each cell test. A thermistor is mounted to the side of each cell as near as possible to the center of the electrode structure. The controllable test parameters are summarized in Table 2.

A magnetic tape recorder is built into the breadboard unit. The recording of data during the cell tests is done in a sampled mode so that 14 s of data are recorded every 72 s. This mode, as well as the total tape capacity, is commensurate with the capabilities of the recorder that will be used in the flight experiment. For test purposes, and for the data presented here, a Sanborn chart recorder is used.

### 3. Discussion and Results

A chart showing part of the data for a typical cell is shown in Fig. 3. This is the second set of tests to which the cell was subjected. The cells were cycled at 100% depth of discharge based on the rated cell capacity. The initial cell capacity was measured to be 6.54 A-h. The final cell capacity was 5.53 A-h after five cycles; this is a reduction of 15.5% during the second set of tests.

The normal behavior of a silver-zinc cell on charge is shown in Fig. 3. In the region between A and B of the chart, the plot of cell voltage versus time shows the characteristic two plateaux. The lower plateau corresponds to the formation of argentous oxide. These features may

also be observed in the record of the charge current and the 2.5-V power supply current.

In the region of the chart from B to C, normal discharge behavior is illustrated. Again, a double plateau profile is shown by the cell voltage. The constant current discharge is shown by the square wave structure of the discharge current record. In addition, it should be noted that, as would be expected, the temperature of the cell rises appreciably during discharge. The highest temperature recorded was 43°C.

The capacity charge is shown between C and D. Except for the longer time required, the features of the plots are the same as during the preceding charge.

The polarization discharge is shown between D and E. The behavior of the cell is similar to that observed during the earlier discharge. Between E and F, the polarization measurement is shown. The polarization voltage, shown by the maximum height of the current ramp at point P, decreased to 1.08 V at a current level of 13.2 A. At this point, the cell voltage was reduced to the low-level cut-off voltage before the current ramp was complete and the polarization test was terminated.

Between F and G, and G and H, the final capacity charge and capacity discharge, respectively, are shown.

One result of the tests performed on the breadboard unit will be the definition of test parameters which will ensure that the cells selected for the flight experiment will show significant degradation in a 1-g environment. Thus, a basis for comparison of flight data and 1-g control data will be established. It may be concluded from the data already obtained that this objective may readily be attained for the 5-A-h cells.

Table 2. Controllable parameter ranges

Parameter	Range
Charge level sensor voltage, V	0.54 to 10
Discharge level sensor voltages, V	0.45 to 2.4
Polarization level sensor voltage, V	0 to 10
Cell rejection level sensor voltage, V	0 to 10
Discharge current, A	0.7 to 20
Polarization discharge current, A	0.7 to 20
Capacity discharge current, A	0.14 to 10
Polarization ramp current, A	0 to 20
Polarization ramp rate, s/full scale current	1 to 1200
Charge, A	0 to 10
Capacity charge, A	0 to 10

## C. Charge Recovery and Open-Circuit Voltage Studies of the Liquid Potassium-Amalgam Electrode, G. L. Juvinall

### 1. Introduction

Some of the unusual properties exhibited by liquid amalgam electrode systems have been described previously in SPS 37-55, Vol. III, pp. 105-107, and SPS 37-57, Vol. III, pp. 88-89. The investigation of the power-producing reactions involved in the operation of liquid amalgam electrodes and their possible applications to

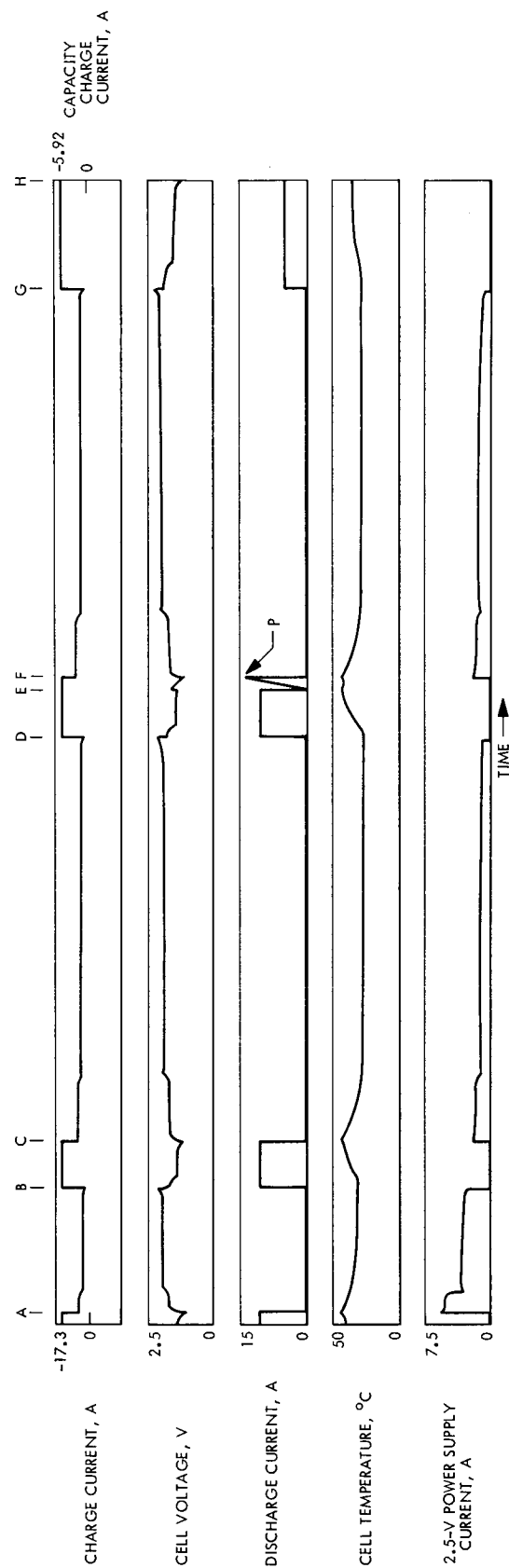


Fig. 3. Results of second test of cell at 100% depth of discharge

space power systems is a major objective of the JPL battery advanced development effort. Dr. G. Myron Arcand is the principal investigator on this project.

Several amalgam electrodes show promise as rechargeable anodes for galvanic cells; all are alkali-metal amalgams. However, most of the electrodes have a fairly high rate of self-discharge. The K(Hg)/KOH (saturated) electrode has a reasonable shelf-life, and may be potentially the most useful of the group. The effects of total charge on recovery and open-circuit voltage of the potassium-amalgam electrode are reported here.

## 2. Experimental Approach

The method for measuring high discharge rates by means of a hanging-drop mercury electrode is described in Ref. 1. The hanging-drop electrode permits the use of very high current densities with normally available equipment. In this way, very small known surface areas can be obtained and small amounts of material are required. The electrodes are discharged in the usual manner and the discharge behavior at high rates is followed with an oscilloscope. At the highest rates, discharge times of less than one second are observed.

The reference electrode is a Hg/HgO electrode with a Luggin capillary mounted as close as possible to the working electrode. The auxiliary electrode is a platinum mesh cylinder.

In order to investigate the effect of total charge on open-circuit voltage and charge recovery, a hanging-drop K(Hg) electrode was operated in saturated KOH at a discharge rate of 2688 mA/cm<sup>2</sup>. A fresh drop was used for each experiment. The total charge was varied and the open-circuit voltage and percent recovery determined for each case.

## 3. Discussion and Results

The effects of total charge (amalgam concentration) on open-circuit voltage ( $E_{oc}$ ) and charge recovery are shown in Table 3. It can be seen that the recovery is poor at the low concentrations and increases steadily with increasing concentration. The recovery stabilizes at 4 or 5 mA-h/g-Hg with only a slow increase up to 9 mA-h/g-Hg.

The open-circuit voltage also decreases steadily with increasing amalgam concentration. This decrease may be anticipated. However, calculations based upon the Nernst equation indicate that a maximum change of 59 mV can

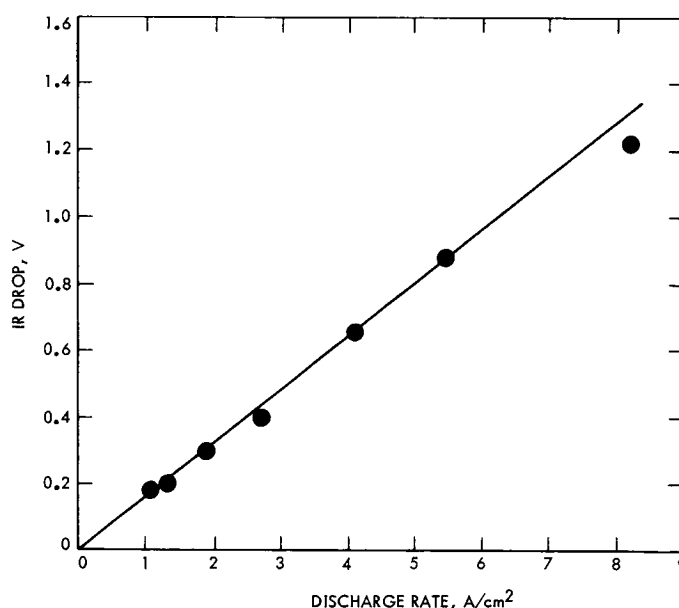
**Table 3. Effect of amalgam concentration on open-circuit voltage and charge recovery<sup>a</sup>**

Concentration, mA-h/g-Hg	$E_{oc}$ vs Hg/HgO, V	Recovery, %
1	-1.62	29
2	-1.64	61
3	-1.66	74
4	-1.69	85
5	-1.70	85
6	-1.73	87
7	-1.74	88
8	-1.75	89
9	-1.76	90

<sup>a</sup>For K(Hg) electrode, discharge rate = 2688 mA/cm<sup>2</sup>; for Hg/HgO electrode,  $E = -0.924$  V vs standard H<sub>2</sub> electrode.

be expected as compared with the observed change of 140 mV. There should be no detectable effect due to non-ideality in the aqueous phase because the concentration of potassium ion remains constant. Thus, non-ideal behavior in the amalgam is indicated.

The effect of discharge rate on  $IR$  drop has also been investigated by means of a hanging-drop system in which the tip of the Luggin capillary is positioned approximately 0.5 cm from the drop. All drops were charged to 9.1 mA-h/g-Hg. The results of these experiments are summarized in Fig. 4. The linear relationship between current density and voltage drop strongly suggests a resistive effect. All data points represent duplicate measurements except for



**Fig. 4. Effect of discharge rate on  $IR$  drop for amalgam electrodes**

the point at 8.15 A/cm<sup>2</sup>. This experiment was conducted to verify that the low voltages observed during high-rate discharges are not due to limitations inherent in the electrode system.

#### Reference

1. Arcand, G. M., *Investigation of Electrode Materials for Alkaline Batteries*, Fifth Quarterly Report, JPL Contract 952265, Idaho State University, Pocatello, Idaho, Aug. 25, 1969.

### D. Solar Cell Research, P. A. Berman

#### 1. Introduction

One of the major concerns in the utilization of silicon solar cells for space is the fact that significant decreases in power can occur when the cells are subjected to irradiation by electrons and protons that exist in the Van Allen Belts surrounding the earth and proton flares that occur in deep space as the result of solar activity. This effect requires the prediction of solar cell performance in various space environments (e.g., various combinations of temperature and solar intensity), and involves improved characterization of solar cell parameters. The objective of this program is to determine the effects of space-type environments such as radiation, temperature, and solar intensity on the operating characteristics of state-of-the-art and developmental-type solar cells.

The work discussed in this article is being carried out by The Naval Research Laboratory (NRL) under JPL direction. The program includes the following major tasks:

- (1) Investigation of low-flux irradiation on experimental lithium diffused silicon solar cells. The objective is to determine the effects of irradiation at rates comparable to those found in near-earth space (as opposed to the very high irradiation rates associated with the utilization of particle accelerators).
- (2) Determination of the electrical characteristics of irradiated and unirradiated solar cells at low temperatures and intensities to establish the applicability of solar cells to outer-planet missions.
- (3) Fundamental investigation of the action of lithium in irradiated silicon to yield a physical model for radiation damage annealing in lithium-doped, silicon solar cells.

The progress and status of each of these tasks are discussed below.

#### 2. Low-Flux Irradiation

Investigations of low-flux gamma irradiation on lithium-diffused silicon solar cells have been initiated. Three sample chambers are utilized for this purpose, each chamber being capable of holding twenty-six 1 × 2-cm solar cells with individual 10-Ω load resistors across each cell. Initially tungsten wire quartz lamps fabricated at the NRL were considered for the illumination source, since no suitable commercial lamps could be found which provided uniform light over the required area and yet fit within the confined space of the radiation chamber. The test run indicated severe degradation of the lamps, and subsequent examination showed a gray film deposited on the inside walls of the lamps. Replacement by two new lamps resulted in a similar failure. Consequently, the use of this type of lamp was discontinued. A set of five 12-V automobile lamps was subsequently investigated as an illumination source. The heat generated caused the solder connections between the lamps to melt; however, the use of argon gas to aid heat conduction from the lamps eliminated this problem and the solar cells were maintained at the desired temperatures of 30 and 60°C.

The cells being investigated in the gamma ray experiments are Heliotek cells having four basic groupings, namely: (1) high lithium concentration in float-zone material, (2) low lithium concentration in float-zone material, (3) high lithium concentration in crucible-grown material, and (4) low lithium concentration in crucible-grown material. A group of 10-Ω-cm state-of-the-art N/P cells is being utilized for comparative purposes. All cells were measured under a Spectrosun X-25L solar simulator calibrated at 140 mW/cm<sup>2</sup> air mass zero by a balloon flight standard cell prior to test. Two of the irradiation chambers are in a gamma field whose strength is  $4.2 \times 10^3$  R per hour, which is equivalent to nearly  $3 \times 10^{12}$  1-MeV electrons/cm per day on an N/P cell. One chamber is in the pool but not in a radiation field for control purposes. One of the state-of-the-art N/P solar cells in each chamber is being monitored daily for its current through a 10-Ω load resistor. The variation in light uniformity on the cell plate is approximately ±10%.

#### 3. Low-Temperature—Low-Intensity Investigations

Work on low-temperature electron irradiation of solar cells has been initiated. Both N/P cells with and without solder coatings are being investigated as well as developmental cells having an additional aluminum P<sup>+</sup> layer at the back surface. Pre- and post-irradiation measurements were carried out at illuminations of 1, 2, 5, 10, 20, 35, and

70 mW/cm<sup>2</sup> obtained with a Spectrosun X-25L solar simulator using neutral density filters. Temperatures were measured by a copper-constantan thermocouple soldered to each cell. The special aluminum-doped cells exhibited no anomalies in current-voltage curve shape prior to irradiation by 1-MeV electrons. The short-circuit current decreased by approximately 10% in going from room temperature to 110°K, while the open-circuit voltage increased by about 35%, the net result being an increase in conversion efficiency from 11.3 to 15.2% as the temperature was decreased. The state-of-the-art non-aluminum-doped cells exhibited a "tail" in the vicinity of the normal open-circuit voltage at 110°K. The 1-MeV electron irradiations were carried out at a temperature of 115°K. After a fluence of 10<sup>15</sup> electrons/cm<sup>2</sup>, the short-circuit currents were reduced to between 40 and 50% of the unirradiated values at a light intensity of 5 mW/cm<sup>2</sup>. It was noted that after a fluence of 10<sup>13</sup> electrons/cm<sup>2</sup> the aluminum-doped cells also developed a "tail" in the current-voltage curves near open-circuit voltage, in spite of the fact that this anomaly did not occur prior to irradiation. The phenomenon grew progressively more pronounced with increasing irradiation fluence.

#### 4. Lithium-Doped Silicon

Further improvements of the Hall measurement techniques utilized in the investigation of lithium-doped silicon are being made. Techniques for mounting the samples are being investigated to assure that sample stresses will not yield spurious results. Calibration of the carbon resistance and platinum resistance thermometers for determination of the temperatures of the samples during the Hall experiments has been carried out against germanium resistance thermometers. The agreement was of the order of a few tenths of 1%.

### E. Evaluation of the Electrical Performance of Mercury-Cadmium Cells, R. E. Patterson and R. S. Bogner

#### 1. Introduction

As missions become more complex and extend for longer durations, the severity of requirements imposed on flight batteries increases. One such requirement on the battery system is that its coulombic efficiency must be maintained after long periods of inactivity. JPL is currently evaluating the suitability of available battery systems for future planetary missions requiring long transit periods. A previous article (SPS 37-60, Vol. III, pp. 73-76) discussed the evaluation of secondary Ag-Zn and Ag-Cd cells. This article discusses the evaluation of charged primary Hg-Cd

cells, based on the mercuric oxide-cadmium electrochemical system that has been reported to have long wet-shelf life capabilities (Ref. 1). The theoretical energy output of this system is 67 W-h/lb.

#### 2. Experimental Design

Sixty primary mercury-cadmium cells were tested according to the outline shown in Fig. 5. The cells were rated by the manufacturer at 20 A-h. Storage tests at temperatures of 75, 120, 160, and 200°F were performed in forced-air, constant-temperature chambers. Upon completing the storage periods indicated in Fig. 5, the cells were discharged at room temperature at a rate of 5.0 A to a cutoff voltage (per cell) of 0.5 V. Performance test discharges were performed at rates of 0.5, 1, 5, and 10 A and at temperatures of 32, 75, 120, and 160°F. Voltage was recorded continuously during discharge.

#### 3. Results

A summary of the storage test data is presented in Table 4. The severity of cell degradation increased with storage temperature and at 200°F total degradation was observed. Storage tests at 75°F are still in progress; however, initial data indicates that at this temperature, cell capacity remains fairly constant for storage periods up to 16 mo.

Results of the cell performance characteristic tests are summarized in Table 5. In general, cell capacity and discharge voltage decreased with increasing discharge rate over the current range of 5 to 10 A. Also, it can readily be seen that the best performing cells were those discharged at 160°F. Even at the low discharge rates, none of the performance test cells delivered the rated 20-A-h capacity.

The recharge tests indicated that the cells could not be recharged satisfactorily.

#### 4. Conclusions

Based on the test results, the following conclusions were made concerning the electrical performance of mercury-cadmium cells of the design tested:

- (1) Cells are not suited for high-temperature storage applications.
- (2) Cells can be stored at room temperature for periods up to at least 16 mo with less than 50% degradation.
- (3) Cells are best suited for low discharge rate applications ( $\leq 1$  A).

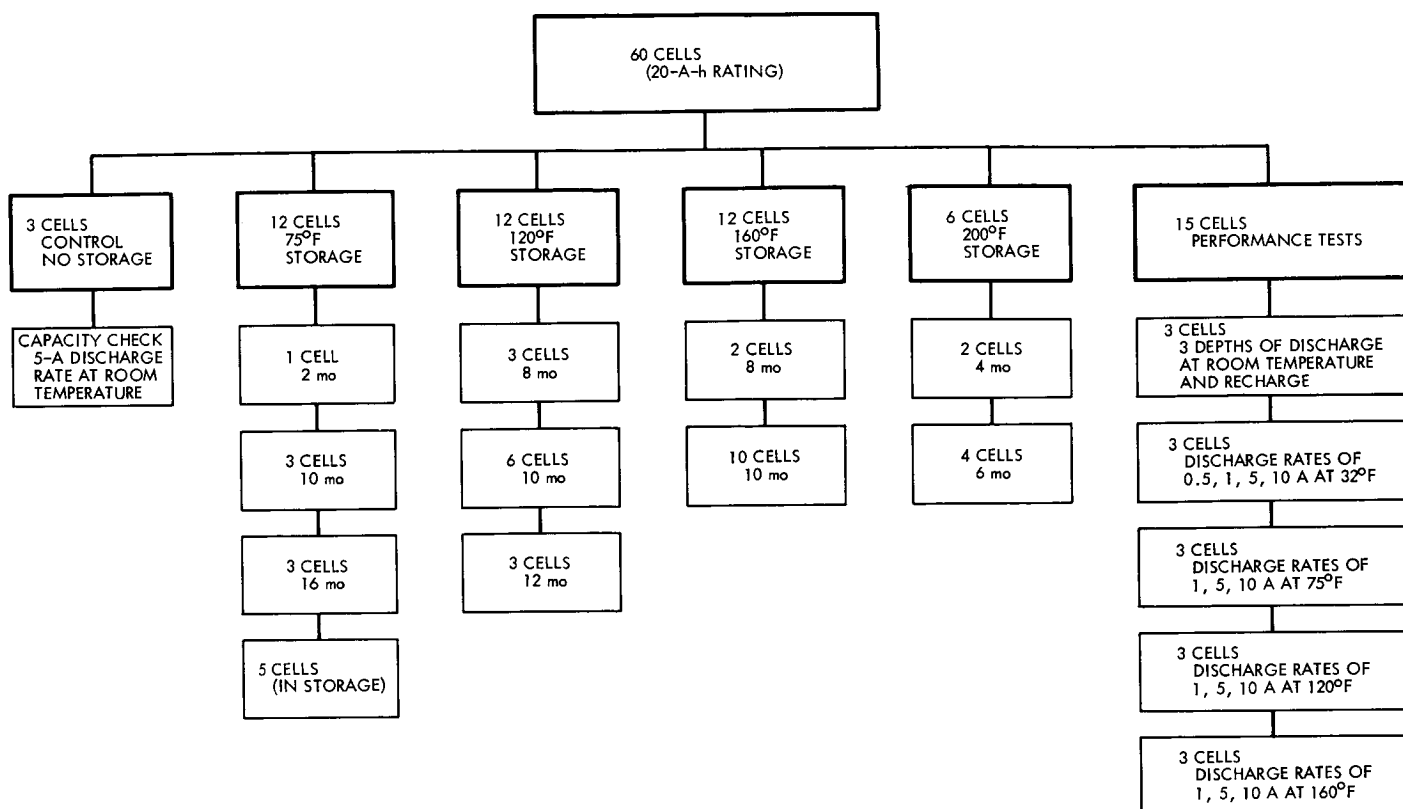


Fig. 5. Mercury-cadmium cell test outline

Table 4. Average cell capacity after storage at selected temperatures for various storage periods<sup>a</sup>

Storage period, mo	Average capacity, A-h			
	75°F storage	120°F storage	160°F storage	200°F storage
2	13.9			
4	10.0			0
6				0
8		10.3	4.5	
10	10.8	7.1	3.9	
12		5.6		
16	11.9			

<sup>a</sup>Cells discharged at 5.0-A rate.

- (4) Cells can be discharged at a temperature of at least 160°F without suffering any decrease in performance.
- (5) Cells are not suitable for recharge applications.
- (6) In general, the performance of the tested mercury-cadmium cells is poor in comparison with both Ni-Cd and Ag-Zn cells.

#### Reference

1. Klein, M. G., and Eisenberg, M., "A New Long Wet-Shelf Life Primary Battery System," *Electrochemical Technology*, Vol. 3, No. 1-2, p. 58, January-February 1965.

Table 5. Capacity-voltage data of HgO-Cd cells

Discharge rate, A	Capacity, A-h				Voltage, V			
	32°F	75°F	120°F	160°F	32°F	75°F	120°F	160°F
0.5	16.6	—	—	—	0.880	—	—	—
1.0	14.4	16.8	14.8	18.8	0.840	0.875	0.900	0.900
5.0	11.6	14.1	10.1	16.2	0.705	0.798	0.815	0.830
10.0	—	10.0	14.0	17.3	—	0.568	0.710	0.755

## F. Electric Propulsion Power Conditioning,

E. N. Costogoe

### 1. Introduction

The electric propulsion power conditioning activities are directed toward accomplishing two tasks. The first task involves the modification of a breadboard power conditioner for operation with a hollow cathode thruster. The second task, performed in conjunction with the work on electric propulsion applications and electric propulsion power conversion, involves the design, fabrication, and evaluation of a complete power conditioning subsystem of the electric propulsion system. The power conditioning subsystem consists of: (1) one breadboard and two experimental power conditioning units, (2) a switching matrix unit that will be utilized to switch the power conditioning units among ion thrusters, and (3) a maximum-power-point monitor to detect the maximum-power point of the photovoltaic power source and determine the power margins. Item (1) is being developed under contract with the Hughes Aircraft Company. All other items are in-house tasks being developed under this work unit. This article will be devoted exclusively to the description of the switching matrix unit.

### 2. Switching Matrix

The switching matrix unit has been designed to interconnect the electrical output of four power conditioners to five ion thrusters and to one test position. The test position is utilized for evaluating the performance of a power conditioner on dummy loads before it is connected to a thruster. All switching takes place when the power conditioners are de-energized.

The block diagram of the switching matrix unit is shown in Fig. 6. The unit consists of three major subassemblies: (1) the logic subassembly, (2) the power switches, and (3) the manual control and display panel (MCDP) subassembly. The switching matrix has two modes of operation: (1) the automatic mode, where all inputs are supplied from the flight controller module to establish the desired interconnection between the power conditioners and ion thrusters; and (2) the manual operation mode, where the flight controller is inoperative and the MCDP issues the commands for the interconnections. The display panel identifies the interconnections of the power conditioner and ion thruster for both modes of operation. Two of the subassemblies, the logic and the power switches, are designed for operation in  $10^{-6}$ -torr vacuum over a temperature range of 0 to 130°F. The MCDP is part of the ground support equipment that is expected to operate in a con-

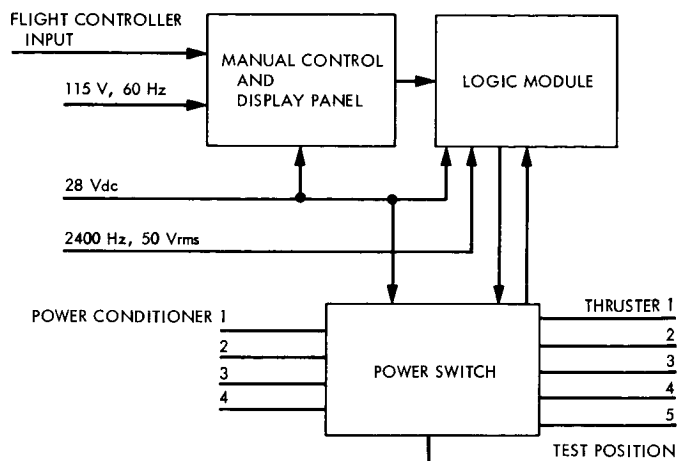


Fig. 6. Switching matrix block diagram

trollable environment of 70 to 80°F temperature. A brief description of the functional characteristics of each subassembly is presented below.

**a. Logic subassembly.** The functional block diagram of the logic subassembly is shown in Fig. 7. The logic subassembly receives binary-coded information from the MCDP module and switch-state information from the power switches. The coded information identifies the desired interconnection between a power conditioner and an ion thruster or test position. The code format is shown in Table 6. The three bits of binary information identifying the desired ion thruster (D, E, F) are supplied to the position sense logic circuit. The information is processed by the position sense logic circuit and is sent to (1) the pulse generator to initiate a drive pulse for the switch driver, and (2) to the wrong command sense logic to verify that the command received can be executed. At the same time three binary bits of information (A, B, C), identifying the desired power conditioner, is supplied to the switch encoder circuit. The information is processed by the encoder and is sent (1) to the switch driver circuit to provide the ground connection for the appropriate switch driver circuit, and (2) to the switches to select the proper switch bank.

Table 6. Binary command format

Power conditioner	Binary bits			Ion thruster	Binary bits		
	A	B	C		D	E	F
1	1	0	0	1	1	0	0
2	0	1	0	2	0	1	0
3	1	1	0	3	1	1	0
4	0	0	1	5	1	0	1
				Test position	0	1	1

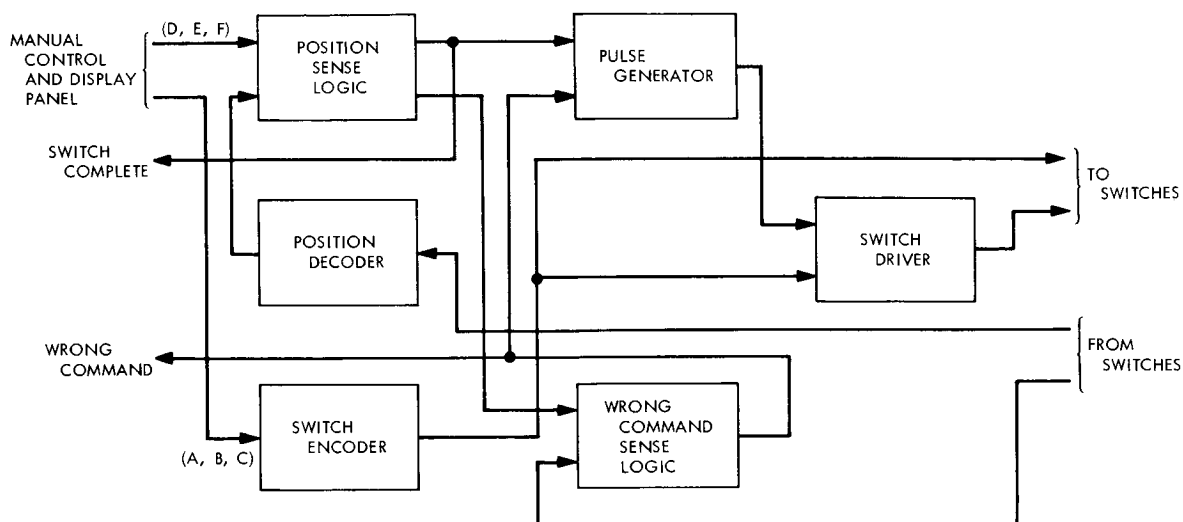


Fig. 7. Logic module functional block diagram

Information from the power switches identifying the existing interconnections of each power conditioner to an ion thruster or test position is supplied to the (1) wrong command sense logic circuit, and (2) to the position decoder circuit. The wrong command sense logic compares the information from the position sense logic and from the switches. If the comparison detects the presence of a wrong command, such as an error in the coded information, or that the elements identified are already interconnected, it generates (1) a signal to identify the presence of the wrong command, and (2) inhibits the pulse generator from issuing the pulse to the switch driver.

The position decoder circuit decodes the status of the switches and sends the decoded information to the position sense logic circuit. This logic circuit compares the decoded status information from the switches with decoded command information received and maintains the pulse drive to the pulse generator. When the position decoder status information and the thruster command information are matched, the pulse drive is terminated and a signal is generated to identify that switching is completed. Interlock circuitry is provided to prevent the following events:

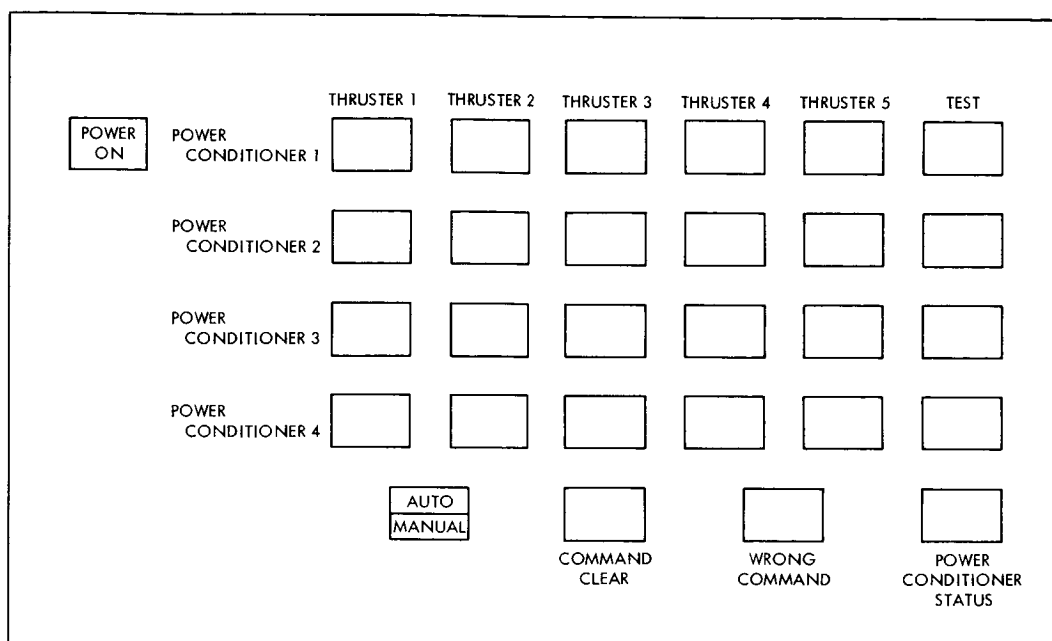
- (1) Connection of one ion thruster to two power conditioners.
- (2) Connection of one power conditioner to two ion thrusters or two power conditioners to the test position.
- (3) Loss of the digital information if the power supply failed temporarily.

**b. Manual control and display panel subassembly.** The manual control and display panel (MCDP) subassembly provides the two modes of operation (the automatic and manual operation modes). In the automatic mode, the unit receives commands from the remotely located flight controller. In the manual operation mode, the commands are generated at the MCDP.

A visual display of the existing connections is provided by the MCDP in both modes of operation. The front panel layout of the unit is shown in Fig. 8. The matrix of illuminated push switches identifies the interconnections between the power conditioners and ion thrusters or test positions. The switches energize the necessary logic circuitry and generate the coded signals to effect the desired interconnection, when the manual control unit is in the manual mode. The WRONG COMMAND light is *on* when the coded signals received identifies a command that cannot be executed because of the presence of an error either in the coded information or existing interconnection. The POWER CONDITIONER STATUS light is *on* when a power conditioner is energized. The signal is supplied to the interlock circuitry to inhibit the switches. The COMMAND CLEAR switch is utilized to clear the wrong command from the logic network.

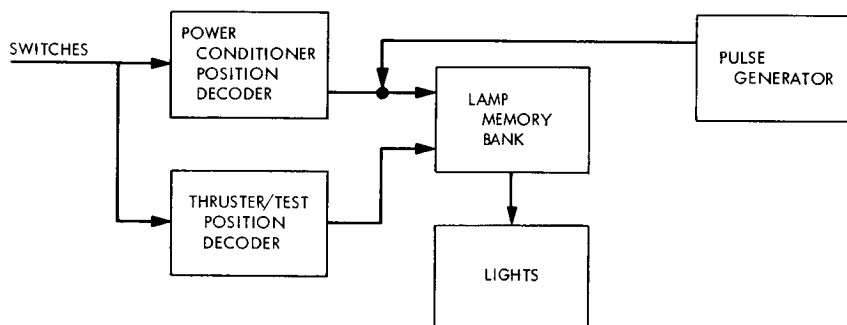
The block diagram of the manual control and display panel is shown in Fig. 9. The manual control unit consists of three circuits.

- (1) The *power conditioner switch and thruster relay circuit* comprise six boards containing four power conditioner switches per board. Each power condi-

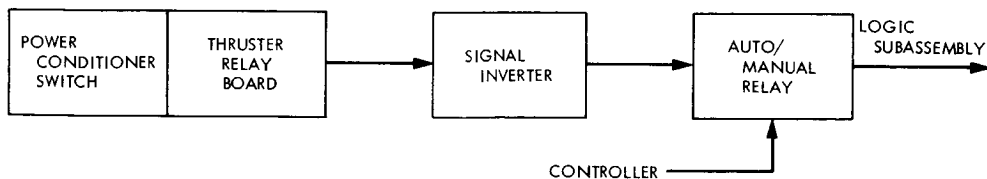


**Fig. 8. Front panel layout**

(a) DISPLAY PANEL UNIT



(b) MANUAL CONTROL UNIT



**Fig. 9. Manual control and display panel block diagram**

tioner switch of each board corresponds to a switch of the matrix shown on the front panel for the selection of the desired interconnection.

- (2) The *signal inverter* is utilized to invert the signal polarity generated by the thruster relay board units for the logic subassembly.

- (3) The *auto/manual relay* selects the signals for either the signal inverter or controller for the desired operation.

The display panel unit receives signals from the switches identifying the interconnection. The signals are decoded

by the power conditioner and thruster/test position decoders. The output of the decoders is supplied to the lamp memory bank unit. A pulse generator is connected to the output lines of the power conditioner decoder. The pulse is transmitted only if the power conditioner position decoder output is of the proper potential. There are six lamp memory bank units corresponding to the six rows of lights for the five thrusters and test position. Appropriate inputs from the thruster/test position decoder and the presence of the pulse for the appropriate power conditioner position decoder energizes the light identifying the interconnection between the power conditioner, thruster, and/or test position.

**c. Power switches.** The switches are Tech Laboratories Inc. hermetically sealed units and have silver alloy contacts to carry 8 A at 2000 Vdc. Each switch has 4 decks, 2 poles per deck and 6 positions per deck. They are solenoid-driven stepping units with non-shortng contacts.

Eight switches are required for the interconnection of the four power conditioners to the five thrusters and one test position. The switches are ganged in groups of two to interconnect each power conditioner to the system. Eight wipers of switch 1 and six of switch 2 are used to transmit power. The remaining wipers of the switches are used for internal control within the switching matrix.

## **G. Long-Term Space Radiation Environment Testing of Lithium-Diffused Silicon Solar Cells, J. M. Weingart**

JPL is responsible for coordinating a program of research and development of the lithium-diffused solar cell, involving university, industrial, and cell manufacturer contracts, as well as in-house research, covering activities from basic research on the nature of radiation damage in lithium-diffused silicon to the fabrication of lithium-diffused silicon cells. This program was originated at the Goddard Space Flight Center and subsequently transferred to JPL in 1968. The organization and various activities of this program have been described in a number of JPL publications (SPS 37-60, Vol. III, pp. 62-63, and Ref. 1).<sup>3</sup> This article presents a program to evaluate the performance of lithium-diffused cells in a simulated space radiation environment.

<sup>3</sup>Also *Semiannual Review of Research and Advanced Development, July 1-December 31, 1969: Vol. II (OART)*, in press (JPL internal document).

This program to investigate lithium cells began after Wysocki (Ref. 2) and others demonstrated that the introduction of lithium by thermal diffusion techniques into P/N silicon solar cells resulted in increased recovery of irradiated cells over the recovery of comparably irradiated nonlithium cells. Certainly one of the most serious problems associated with photovoltaic arrays has been power degradation due to the space radiation environment, which consists primarily (from the point of view of cell damage) of the belts of trapped electrons and protons in near space and protons from solar flares in deep space. Electrons generate primarily point defects in silicon; protons cause much more massive damage, such as defect clusters, and loop and spiral dislocations on a macroscopic scale.<sup>4</sup> All of these defects introduce states in the forbidden band of silicon and act as trapping and recombination centers for minority carriers. The degradation of minority carrier diffusion length in conventional solar cells results in fewer minority carriers being swept across the P-N junction for a given minority carrier injection level. The result is a decrease in power conversion efficiency.

The presence of lithium in irradiated silicon results in a change in the apparent response of the material to irradiation. The physical damage appears to be independent of the presence of lithium. However, because of the extremely high diffusion coefficient (Refs. 3 and 4) of lithium in silicon, and because a considerable percentage of the diffused lithium remains electrically active (unlike other materials like copper, for example, which has a similar diffusion coefficient in silicon but a strong predilection for precipitation at defect and dislocation sites), a considerable percentage of lithium remains "in solution" until encountering a trapping or recombination center. There it can form new complexes with existing vacancies, divacancies, and other complexes. The effect of the participation of lithium in the formation of these complexes is a net reduction in the recombination cross section for electron-hole pair recombination.

It was originally hoped that the introduction of lithium into silicon solar cells would be a simple and reliable method for increasing the radiation resistance of the cells. During the course of more detailed investigations (Footnote 3), a number of significant complications became evident, including the tendency of lithium cells subjected to irradiation to redegrade after apparent par-

<sup>4</sup>Weingart, J., and deWys, E. C., article on *X-Ray Topographic Analysis of 10-MeV Proton Damage in Silicon Solar Cells* (in preparation, title not final).

tial recovery from damage, and electromechanical degradation of the electrical contacts over long periods of time. The research and development program has addressed itself to understanding these problems, as well as understanding the fundamental nature and behavior of the various complexes formed in irradiated lithium-diffused silicon and silicon cells, with the objective of developing reliable, practical radiation-hardened cells.

While we still do not have a complete understanding of the nature and behavior of these defects and complexes, and do not yet have a full model to explain the changes in the electrical characteristics of the actual cells in terms of the behavior of these defects and complexes, it is nevertheless important to evaluate in a pragmatic way the behavior of the current state-of-the-art lithium cell through long-term exposure to a simulated space radiation environment. Monoenergetic electron and proton irradiation of cells has been a useful technique to study radiation damage in cells and the resulting changes in electrical parameters of the cells. However, this technique can only be reliably useful in evaluating the potential flight characteristics of cells after there has been careful correlation with long-term simulated environmental tests. In fact, the usefulness of a 1-MeV electron damage equivalence with damage in a realistic environment can be determined only by such a correlation. The concept of a meaningful 1-MeV damage coefficient to predict electron radiation degradation under actual flight conditions by comparison with monochromatic irradiation has in fact come under serious criticism (Ref. 5); the program described below will be helpful in evaluating the range of usefulness of such a damage coefficient. Hopefully the synergistic effect of concurrent research, device development, and device testing will lead in the most efficient way to a practical radiation-hardened cell for future flight missions.

Two industrial contractors have been chosen to study the real-time effects of a simulated space radiation environment on illuminated and dark lithium-diffused and (for comparison) conventional N/P *Mariner* cells. This work is being carried out by Philco-Ford Corporation, Space and Re-entry Systems Division, Palo Alto, California (JPL contract 952585), and by Lockheed Aircraft Corporation, Lockheed-Georgia Company Division, Marietta, Georgia (JPL contract 952586).

The tests consist of subjecting a large group of cells, representing state-of-the-art lithium-diffused cells and conventional N/P cells, to a low-flux spectral distribution of electrons in an organic vapor-free vacuum chamber. Groups of cells will be temperature-regulated at four

discrete temperatures over the range  $-50$  to  $+80^{\circ}\text{C}$ , with all of the cells loaded near their pre-irradiation maximum power points under air mass zero (AM0) one-sun conditions. During the course of irradiation most of the cells will be continuously illuminated, either by water-filtered tungsten illumination (Philco) or blue-rich xenon illumination (Lockheed). Some of the cells will remain dark during the entire run to determine the effects, if any, of illumination on the radiation degradation characteristics of the cells. Distribution of experimental parameters is shown in the experimental matrices (Tables 7 and 8).

While the contractors have approached the program in somewhat different ways, there has been standardization of key aspects of the tests. Strontium-90 beta sources, having a beta spectrum similar to that encountered in near space, will provide the low-flux rate (approximately  $10^{11}$  electrons/cm<sup>2</sup>-s) irradiation. Both systems will employ sorption and vac-ion pumps to provide an organic vapor-free vacuum to eliminate any effects of hydrocarbon contamination, and will incorporate 1-in.-thick Corning 7940 fused silica windows to permit continuous illumination of the cells during irradiation as well as periodic current-voltage (*I-V*) characteristic measurements. Corning 7940 is among the most resistant to electron and gamma irradiation of the available optical materials having the required UV transmission characteristics. Incorporation into both systems insures that there will be

Table 7. Lockheed test matrix<sup>a</sup>

Type of cell	$-50^{\circ}\text{C}$		$30^{\circ}\text{C}$		$60^{\circ}\text{C}$		$80^{\circ}\text{C}$	
	Light	Dark	Light	Dark	Light	Dark	Light	Dark
Without cover glass								
Float-zone silicon:								
Low dope			5		5		5	
High dope	5		5		5	5	5	
Pulled silicon:								
Low dope			5	5	5	5	5	
High dope	5		5	5	5	5	5	
Standard N on P cell	3		3	1	3	1	3	
With cover glass								
Float-zone silicon:								
Low dope					5			
High dope								
Pulled silicon:								
Low dope					5			
High dope	5		5	5	5		5	

<sup>a</sup>All tests performed in vacuum of  $<10^{-6}$  torr; data taken *in situ* at fluences of  $10^{12}$ ,  $2 \times 10^{12}$ ,  $5 \times 10^{12}$ ,  $7 \times 10^{12}$ ,  $10^{13}$ ,  $2 \times 10^{13}$ ,  $5 \times 10^{13}$ ,  $7 \times 10^{13}$ ,  $10^{14}$ , and  $2 \times 10^{14}$  electrons/cm<sup>2</sup>; radiation rate of  $10^{12}$  electrons/cm<sup>2</sup>/day.

Table 8. Philco test matrix

Type cell	Diffusion characteristics <sup>a</sup>	Manufacturer	Number available <sup>b</sup>	Strontium-90						1-MeV accelerator, dark, unloaded, 20°C	Spare
				Illuminated and loaded					Dark, unloaded		
				-50°C	20°C	50°C	80°C	20°C	20°C		
Lithium-diffused crucible-grown silicon:											
A	425-90-60	Heliotek	30	4	4	4	4		4	4	6
B	425-90-60	Centralab	20	4	4		4			4	4
C	450-20-0	Centralab	20	4	4		4			4	4
D	425-90-60 <sup>c</sup>	Heliotek	12		4					4	4
Lithium-diffused P/N float-zone silicon:											
E	425-90-120	Heliotek	30	4	4	4	4		4	4	6
F	350-90-60	Heliotek	20	4	4		4			4	4
G	Boron 1	Centralab	20	4	4		4			4	4
H	Boron 2	Centralab	20	4	4		4			4	4
N/P:											
I	10-Ω-cm conventional	Centralab	28	4	4			4	4	4	8
J	10-Ω-cm conventional	Heliotek	10				4	4		2	
K	10-Ω-cm old	Hoffman						4		4	
L	10-Ω-cm pre-irradiated	Hoffman						2		2	
M	10-Ω-cm with cover slide							4		4	
Maximum cell capacity at temperature-regulated blocks				36	36	10	36	18	14		
<sup>a</sup> Diffusion temperature is in °C; diffusion time is in minutes. <sup>b</sup> Cells A-J purchased under present contract; cells K-M are contractor-owned cells from previous contracts included for comparison. <sup>c</sup> 1-mil integral cover glass.											

minimal differences in the optical characteristics of the two systems. In addition, both contractors will use Spectrolab X-25 solar simulators to make the *I-V* cell measurements. JPL has imposed strict requirements on the intensity profile, absolute intensity, and spectral distribution to insure reliability and reproducibility of the data over the duration of the test, and to permit meaningful comparison with any subsequent lithium cell test programs. These procedures will include periodic measurements on the same sets of secondary standard cells by JPL and both contractors. One of the difficulties with previous lithium cell test programs has been the lack of suitable standardization of test environments and *I-V* measurement procedures, as well as the lack of adequate statistics, making it extremely difficult to correlate the results of different programs or to compare the results with subsequent programs. We intend to eliminate such uncertainties and variations in this program.

The electrical characteristics of the cells will be measured periodically during the 6-mo irradiation period. Approximately 12 to 15 such sets of data will be acquired. A data acquisition system will scan the array of cells,

measuring the electrical characteristics of each. In addition to the characteristics in the illuminated power (fourth) quadrant, the forward dark diode characteristics and the dark reverse current characteristics will be measured (first and third quadrants of the *I-V* plane). Changes in these characteristics will also be correlated with electron radiation damage.

Conversion of the data to punched cards for automatic data processing will facilitate computer analysis. Changes in the overall characteristics of the cells during the irradiation period will be plotted, along with changes in the short-circuit current, open-circuit voltage, series resistance, and maximum power as a function of time (fluence) for each class of cells. Statistical analysis (*f* and *t* tests) of the data will determine the significance of any conclusions drawn from the results of the tests, and will provide a measure of the statistical reliability of the averaged data. Duplicate sets of data in digital form will be sent periodically to JPL for incorporation into a central computer facility being designed to correlate environmental tests on lithium-diffused cells and conventional cells tested under the lithium cell program. Data from

industrial, university, manufacturer, and in-house programs will be correlated and studied using this facility.

At the end of the 6-mo irradiation period, the low-temperature cells will be annealed at room temperature and the change in electrical properties measured. Selected lithium-diffused cells will have their contacts removed and redeposited by the manufacturer and their characteristics remeasured, to separate electromechanical contact degradation from radiation damage degradation. Some cells from each group will be stored at room temperature to determine their shelf-life characteristics.

Long-term irradiation is expected to begin by March 1970 on both programs. Hardware construction and assembly is in the final phase and results of the irradiation tests will be reported in a future SPS.

### References

1. *Proceedings of the Conference on Effects of Lithium Doping on Silicon Solar Cells*, Technical Memorandum 33-435. Edited by P. A. Berman. Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1969.
2. Wysocki, J., et al., *Radiation Damage in Silicon*, Final Report to Goddard Space Flight Center, Contract NAS 5-3788. RCA Laboratories, Princeton, N. J., 1964.
3. Fuller, C. S., and Ditzenberger, J. A., *Phys. Rev.*, Vol. 91, p. 193, 1953.
4. Fuller, C. S., and Severiens, J. C., *Phys. Rev.*, Vol. 92, p. 1322, 1953.
5. Stroud, R. H., and Barrett, M. J., *An Analytical Review of the ATS-1 Solar Cell Experiment*, Final Report to Goddard Space Flight Center, Contract NAS 5-11663. Exotech Inc., Washington, Nov. 20, 1969.

## H. Reference Design Criteria for the Thermoelectric Outer-Planet Spacecraft Radioisotope Thermoelectric Generator Power Source,

O. S. Merrill

### 1. Introduction

At present there exists no known power source, radioisotope thermoelectric generator (RTG) or otherwise, that will satisfy the performance, safety, lifetime, and power requirements for an extended outer-planet mission, such as the Grand Tour mission—the reference mission for the Thermoelectric Outer-Planet Spacecraft (TOPS) Project. It is anticipated that at least a 1.5- to 2-yr program is

required to develop useful prototype units. The mechanical, thermal, electrical, magnetic, and nuclear radiation interfaces between an RTG power source and the spacecraft are heavily dependent on the nature and design of the power source. It is felt that the best design of the RTG power source will be one that is tailored specifically to these types of missions.

This article presents the reference design criteria for the development of a radioisotope thermoelectric generator for the NASA-JPL Thermoelectric Outer-Planet Spacecraft Project.

### 2. Functional Description

The function of the TOPS power source is to provide a central supply of electrical power to operate the electrical equipment on board the spacecraft. Power will be furnished by three (or multiple) RTGs electrically connected in parallel and mounted in tandem, or parallel, on a deployable support boom.

The RTG power source will generate electricity at a constant voltage using silicon-germanium thermoelectric elements heated by the decaying radioisotope plutonium 238.

Assuming three RTGs in parallel (electrically), and some small redundancy in the number of thermoelectric couples per RTG, the reliability per RTG will be 0.98 for a 12-yr operating life. This does not include consideration of the probability of environmental survival.

### 3. Interface Definitions

*a. Mechanical.* Two methods of mechanically mounting the RTG power source to the support boom will be considered:

- (1) Mounting the RTGs in tandem.
- (2) Mounting each RTG independently to a "cage" or support rods extending from the boom along the length of the RTGs and external to them.

Each method will be considered for weight, effect on design and performance of RTGs, etc., and compared. Particular attention shall be paid to the dynamic environmental specifications contained in JPL document "Environmental Requirements."<sup>5</sup>

<sup>5</sup>"Environmental Requirements," in *Thermoelectric Outer Planet Spacecraft Functional Requirements* (JPL internal document).

**b. Electrical.** The power source will interface electrically only with the power-conditioning equipment. Together with whatever ancillary equipment is required, they will constitute the power subsystem.<sup>6</sup>

**c. Thermal.** A conductive thermal interface will exist between the power source and its support structure. Radiant thermal interfaces will exist between the power source and all other subsystems of the spacecraft lying within its field of view.

Consideration will be given to the utilization of a portion of the waste heat from the RTGs to heat the science instruments, electronics, and other portions of the spacecraft in lieu of electrical heating.

**d. Magnetic.** A magnetic interface will exist due to the generation and flow of current in and from the power source.

**e. Nuclear radiation.** The power source emits gamma rays and neutrons of sufficient quantity and intensity so as to create potentially severe interfaces with the scientific experiments, the spacecraft electronics, and possibly with other subsystems. The gamma and neutron radiation flux from the RTG will be accurately determined and experiments performed to assure the compatibility of the power source with the sensitive detectors and components of the scientific experiments, spacecraft electronics, and other sensitive or potentially sensitive components.

#### 4. Performance Parameters

**a. Power level.** The maximum power required during the mission occurs at each of the planetary encounters, the last of which, Neptune, occurs near the 11.5-yr end-of-life (EOL). For design purposes, a 12-yr lifetime, a 25% overall degradation in power, and an overall beginning-of-life (BOL) power source conversion efficiency of 5% have been assumed. The unconditioned power requirement at far encounter is 439 W. For design purposes, a parametric power range of  $425 \pm 75$  W electrical, EOL, will be considered.

The power source shall be capable of delivering, while operating in air during the ground checkout and launch sequence, no less than the 248.5 W specified for the launch

phase in JPL document "Power Profile and Allocation."<sup>7</sup> It shall also be capable of providing the power specified during all other mission phases defined in the same document and consistent with JPL document "Flight Sequence."<sup>8</sup>

**b. Lifetime.** The plutonium 238 radioisotope heat sources for the generators will have a shelf-life capability up to 2 yr. The fueled generator will have the capability for a 12-yr lifetime after launch.

**c. Output voltage (EOL).** The design goal for the power source is a nominal, constant output voltage of 30 Vdc.

**d. Specific performance (EOL).** The design goal for the power source alone is 2.0 W/lb, or better, not including other portions of the power subsystem. Current design considerations are based on a 1.7-W/lb figure.

#### 5. Physical Characteristics and Constraints

The power source shall conform to all applicable sections of JPL document "Spacecraft Characteristics and Restraints."<sup>9</sup>

**a. Weight.** The design goal for the weight of the power source only, excluding any support structure or other portions of the power subsystem, shall be  $212.5 \pm 37.5$  lb, consistent with the  $425 \pm 75$ -W (EOL) power requirement, i.e., 2.0 W/lb.

**b. Size.** The volume envelope available for the RTG power source is that defined by a right circular cylinder not exceeding 18 in. in diameter and 72 in. in length.

Should more than three RTGs be required, consideration should be given to at least two alternate RTG arrangements:

- (1) Mounting four in tandem.
- (2) Mounting two sets of two in tandem, but supported from a branching single deployable support boom from the spacecraft.

<sup>7</sup>"Power Profile and Allocation," in *Thermoelectric Outer Planet Spacecraft Functional Requirements* (JPL internal document).

<sup>8</sup>"Flight Sequence," in *Thermoelectric Outer Planet Spacecraft Functional Requirements* (JPL internal document).

<sup>9</sup>"Spacecraft Characteristics and Restraints," in *Thermoelectric Outer Planet Spacecraft Functional Requirements* (JPL internal document).

<sup>6</sup>"Power Subsystem," in *Thermoelectric Outer Planet Spacecraft Functional Requirements* (JPL internal document).

Consideration of an alternate volume envelope may be possible, and should be made in an attempt to determine an optimum number and arrangement of RTGs within the bounds of the spacecraft constraints and the space available under the shroud.

**c. Excess heat rejection mode.** Radiative heat transfer to space shall be the mode of excess heat rejection.

**d. Magnetic field.** The maximum permissible magnetic field created by the RTG power source after injection of the spacecraft into orbit is 0.01 gamma at 10 ft.

**e. Number of RTG units.** The number of RTGs to be used will be determined on the basis of maximum specific power, spacecraft configuration, minimum radiation, and other factors. Particular emphasis is to be placed on the versatility and adaptability of the RTGs for other power levels, other arrangements, and possibly for spacecraft other than TOPS. For design purposes, the baseline TOPS configuration assumes three RTGs mounted in tandem.<sup>10</sup>

**f. Instrumentation.** Sufficient instrumentation shall be provided for each RTG, and for the entire power source, to determine their operating characteristics and to diagnose any abnormal or degrading performance. Experience gained from previous programs (e.g., *Nimbus*-SNAP-19 and ALSEP-SNAP-27) shall be applied to benefit this project.

**g. Radioisotope fuel.** Plutonium 238 will be the fuel used, the form to be selected by the Atomic Energy Commission (AEC), based on extensive evaluation, development, comparison, and the latest and most promising technology suitable to the TOPS requirements. Maximum concern will be exercised in the reduction or elimination of certain light element impurities (oxygen 18, boron, fluorine) and plutonium 236 in order to reduce the number of neutrons from the alpha-*n* reaction and the 2.6-MeV gamma from thallium 208, daughter product of plutonium 236, respectively.

**h. Fuel capsule.** The type and design of the fuel capsule is to be determined by, or under the auspices of, the AEC based on the latest technology, and commensurate with the design of the RTGs.

The length to diameter (*L/D*) ratio of each fuel capsule shall be as large as is practical in order that a minimum cross-sectional area (or "end-on" view) of the capsule closest to the spacecraft electronics bus, and hence also closest to the science instruments, will be exposed. This affords maximum self-shielding by the fuel capsules as well as minimum apparent source size, resulting in minimum unshielded neutron and gamma radiation along that axis.

**i. Environment.** The power source shall conform to all applicable sections of JPL document "Environmental Requirements" (Footnote 5).

## 6. Safety Considerations

The RTG power source, and in particular the fuel capsules within the power source, will be of such a design and have such operational characteristics as to be in compliance with the safety criteria for all phases of the pre-launch, launch, and post-launch sequence, as required by the AEC, NASA, DOD, and the Space Council.

An unofficial document,<sup>11</sup> containing material that the AEC's Multi-Hundred Watt RTG contractor will require for the preparation of the Preliminary Safety Analysis Report, has been prepared. This document is required by the AEC in its Phase I study, and consists of three principal documents: (1) Reference Design Document, (2) Accident Model Document, and (3) Nuclear Safety Analysis Document. Additional or revised information will be furnished to the AEC as required, when available, in order that the contractor will have the most current inputs available for the safety documentation.

## 7. Concluding Remarks

The preceding criteria have been furnished to the Atomic Energy Commission and are currently being used in Phase I of their Multi-Hundred Watt RTG Development Program, under contract with the General Electric Company, Space Division, King of Prussia, Pa. The principal objective of this program is to develop a highly efficient, lightweight, long-life RTG power system for use on a number of potential using agencies' space missions. Among these are the JPL TOPS Project and the planned Grand Tour-type outer-planet missions.

<sup>10</sup>Thermoelectric Outer Planet Spacecraft, Configuration 12J, Sept. 24, 1969 (JPL drawing).

<sup>11</sup>JPL Inputs to the TOPS Preliminary Reference Design Document, Oct. 7, 1969 (JPL internal document); also see following article (Section I) by W. B. Weber.

## **I. Documentation Requirements for Obtaining Launch Approval of an Aerospace Nuclear System, W. B. Weber**

The use of radioisotopes in space is closely regulated and controlled. Special emphasis is placed on the Aerospace Nuclear Safety (ANS) qualifications. Before any nuclear system can be used in space, an unbiased Safety Evaluation Panel (SEP) performs a review with respect to ANS. Although it is the user's responsibility to obtain the launch approval recommendation from the SEP, the Atomic Energy Commission (AEC) has developed a standardized approach to the documentation necessary to support the request for launch approval.

The radioisotope thermoelectric generator (RTG) contractor for the AEC is responsible for developing a Reference Design Document (RDD), an Accident Model Document (AMD), and a Safety Analysis Report (SAR). They are developed in the order listed because they are sequentially dependent. The RTG user (JPL in this case) generally possesses extensive system and mission information insofar as its program is concerned. It is a responsibility of the user to provide the AEC's contractor with this essential input to the RDD. Such is the requirement on TOPS.

The objective of the RDD, the first of the sequence to be generated, is to establish in one document a currently valid and consistent description of all facilities, systems, and operating profiles necessary for accident and safety analyses.

The information required by the RTG contractor has been compiled for the TOPS program according to an outline developed by the AEC for inputs to the RDD.

Much of the information submitted for the preliminary RDD will be typically incomplete and/or not firm at the early stages of most programs. Such status does not relieve the demand for prompt evaluation of ANS qualifications.

The AEC provides that the information in the RDD (and subsequently the AMD and SAR) will be completed and updated by the simple technique of requiring "interim" and "final" versions of all three documents. The timing of the subsequent issues is based on an optimum elapsed-time/time-to-launch schedule that will allow maximum documentation of ANS acceptability. This is intended to assure launch approval recommendation at the earliest practical point in most program schedules.

## IX. Spacecraft Control

### GUIDANCE AND CONTROL DIVISION

#### **A. Thermoelectric Outer-Planet Spacecraft Attitude-Control System Simulation Instrumentation, L. S. Smith**

##### **1. Introduction**

This article presents the relevant features of designs being developed to instrument and actuate a digital attitude-control system that will be mounted on a single-axis air-bearing table. This single-axis simulation of a digital spacecraft system will be adaptable to anticipated changes in control philosophies.

To date a representative digital control technique has been established using as system elements an 11-bit sun sensor, a millidegree-resolution gyro, one momentum wheel, and a hydrazine thruster. With these devices, an interactive design is being established whereby console control and attitude-control functions can be monitored and recorded throughout a testing cycle.

The console-table interaction criteria have been established, and 60% of the detailed logic design for the table has been completed with approximately 20% of the hardware fabricated. Console design will be initiated in the next recording period.

##### **2. System Data Handling Description**

Figure 1 presents a block diagram of the attitude-control system instrumentation system. Three radio link data channels exist between the console and table. Telemetry of table and console data uses two channels with the third channel establishing synchronization between the table and console. There are 7 meaningful words being transmitted serially from the console and 16 words from the table. Digital values of 19 bits can be handled. Spare words are available such that a total of 32 words can be generated at the console and 31 words at the table.

The Word Routing Control block on the console side of Fig. 1 sequentially routes console data into the Console Storage Registers. From these registers the operator can select and display data to verify his setup via the Console Word Display Selector, Display Storage, and Display circuits. Console word formats for table or recorder disposition are arranged within the Data Format Control. Console words are created in binary-coded decimal (BCD) serial formats that must be arranged in binary format for the table and parallel format for the recorder.

The Table Data Receiver data is in either discrete format for status indications or binary format for digital

values. This data is continuously processed by the Input Format Control to convert the binary data to the BCD format. Four table words may be selected by the operator for display via the Table Word Display Selector. Displays are updated 10 times per second.

The data will be recorded on four tracks of a seven-channel recorder in BCD format. The table words are alternately recorded with console words.

Sequencing the above interlacing of data requires synchronization between the console and table functions. Also, to preclude logic race problems, data transfers are phased such that data is shifted at the transmitters prior to being read at the receivers. This phasing is accomplished by using a basic clock frequency four times the data transfer rate to establish grey coded signals at the data transfer rate of 6400 bits/s. Synchronization is then established in the console Clock Syncing Control by

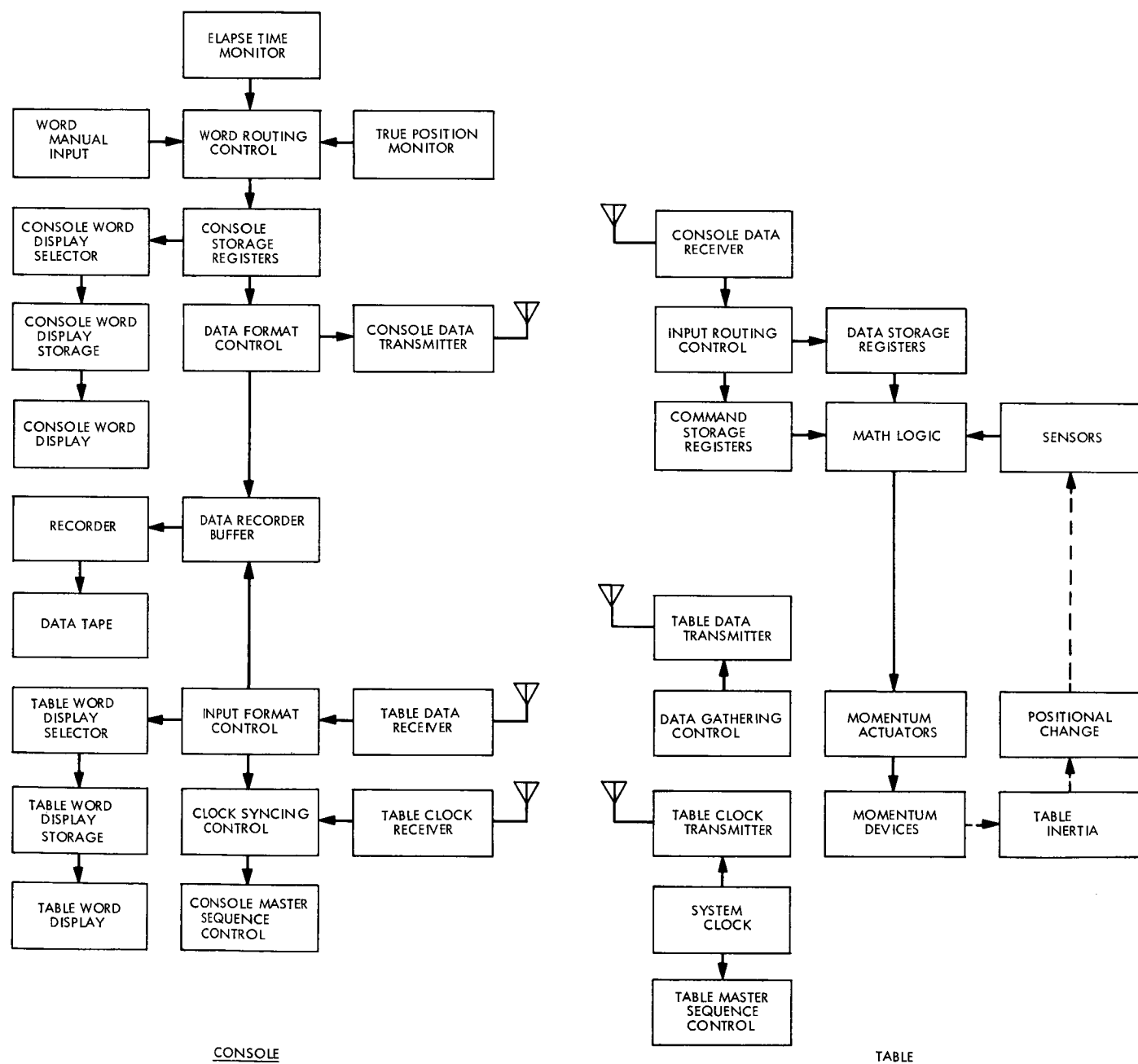


Fig. 1. Single-axis attitude-control simulation data handling block diagram

digitally comparing the table data with a divide-by-four grey counting of the third telemetering channel 25.6-KHz signal to obtain the proper phasing.

The table side of Fig. 1 has an Input Routing Control that places commands in the Command Storage Registers and binary digital values in the Data Storage Registers. Subsequently, depending on the command mode, sensor inputs and console data are used to generate the various attitude-control signals.

Not shown in Fig. 1 are the numerous paths between the Console Master Sequence Control and Table Master Sequence Control blocks and the other respective blocks. These sequencing blocks control the actions described above. Also, the interconnections to the Data Gathering Control block on the table are not shown. This block is integrally associated, for monitoring purposes, with all the table functions.

### 3. Discussion

Additional capabilities are being designed into the simulation system. A shifting of the attitude-control bits can be caused by the console operator to simulate the effects of a variable solar image size. The operator may also control a time-saving simulation of momentum wheel drive power.

The millidegree resolution of the gyro sensors will impose too tight a dead-band restriction on the attitude control during an inertial mode. Currently the proper scaling factor is unknown and can best be determined during the table simulation tests. Consequently, a console control word is available that allows the selection of a digital scaling factor to establish the most effective system.

The size of the currently planned table surface (36 in. in diameter) imposes no stringent requirements on the physical configuration or placement of components making up the attitude-control and instrumentation systems. The only constraints are the leverage arm required for the hydrazine thruster and centering of the optical attitude-control positional sensors.

## B. Analysis of Flexible Structure Effects on Attitude-Control Sensors, E. L. Marsh

### 1. Introduction

A preliminary design version of the Thermoelectric Outer-Planet Spacecraft (TOPS) vehicle is shown in Fig. 2. Observe that the Canopus tracker is attached to the essentially rigid main body of the spacecraft. The Jupiter orbiter version envisions the Canopus tracker at the tip

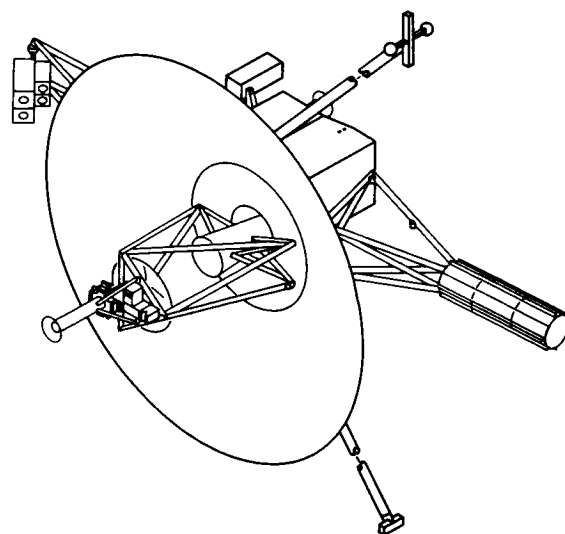
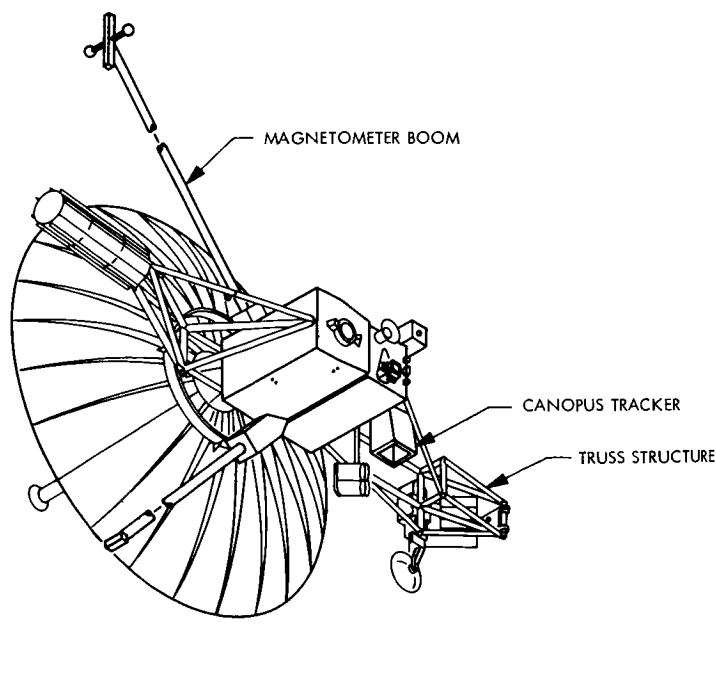


Fig. 2. Thermoelectric Outer-Planet Spacecraft

of the truss structure, in the vicinity of the scan platform. When attached to the main body, the tracker provides correct information regarding the main body's attitude in space; however, flexing of the truss structure can mean that, as in the second design, the tracker can produce information that does not truly determine the main body's orientation in space. The objective of this study is to ascertain the effect of truss flexibility on the sensing accuracy of the Canopus tracker for the TOPS orbiting vehicle.

## 2. Vehicle Dynamics

A convenient scheme for modeling a structure consisting of rigid and elastic components is presented in Ref. 1. The equations used herein are modified forms of those developed in that report, and the reader is directed to it for additional detail.

The basic idea in Likins' work is the approximation of a continuous appendage to a collection of finite rigid bodies that are connected by massless elastic members. It is assumed that the theory of linear elasticity applies to the elastic members.

Each sub-body has six possible degrees of freedom.  $u_i^j$  and  $\beta_i^j$  are the translational and rotational coordinates, respectively, of the  $j$ th sub-body, and for the  $i$ th direction. The matrix  $q$  of deformation coordinates has the form

$$[u_1^1 u_2^1 u_3^1 \beta_1^1 \beta_2^1 \beta_3^1 \cdots u_1^n u_2^n u_3^n \beta_1^n \beta_2^n \beta_3^n]^T$$

In the sequel, modal coordinates  $\eta$  will be introduced for the purpose of decoupling equations.  $\eta$  and  $q$  are related by  $q = \phi\eta$ , where  $\phi$  is a matrix of mode shapes.

For the special case when there are no resultant torques exerted on the flexible appendages, and when all three momentum wheels are activated, the equations of motion for the appendages, total vehicle, and momentum wheels are, respectively,

$$\ddot{\eta} + 2\bar{\zeta}\bar{\sigma}\dot{\eta} + \bar{\sigma}^2\eta = \bar{\delta}\ddot{\theta} \quad (1)$$

$$T = I\ddot{\theta} + J\ddot{\chi} - J\Omega^0\dot{\theta} - \bar{\delta}^T\ddot{\eta} \quad (2)$$

$$\tau = J(\ddot{\theta} + \ddot{\chi}) \quad (3)$$

and where

$$\delta = -\bar{\phi}^T M (\Sigma_{0\epsilon} - \Sigma_{\epsilon 0} \bar{R} - r \Sigma_{\epsilon 0}) \quad (4)$$

The symbols used in the preceding equations are defined as follows:

$\bar{\eta}$  An  $N \times 1$  matrix of truncated modal coordinates, where  $N$  is the number of coordinates retained

$\bar{\zeta}$  An  $N \times N$  diagonal matrix with  $\zeta_i$  representing the percentage of critical damping for the  $i$ th mode

$\bar{\sigma}$  An  $N \times N$  diagonal matrix, where  $\sigma_i$  is the natural frequency for the  $i$ th mode

$\bar{\sigma}^2$  An  $N \times N$  diagonal matrix of squares of natural frequencies

$\phi$  A  $6n \times N$  matrix of eigenvectors, where  $n$  is the maximum possible number of degrees of freedom for the flexible appendages

$M$  The  $6n \times 6n$  mass-inertia matrix for the flexible appendages

$\Sigma_{0\epsilon}$   $6n \times 3$  Boolean operator matrix  $[0\epsilon 0\epsilon \cdots 0\epsilon]^T$ , where  $\epsilon$  and  $0$  are  $3 \times 3$  identity and null matrices, respectively

$\Sigma_{\epsilon 0}$   $6n \times 3$  Boolean operator matrix  $[\epsilon 0 \epsilon 0 \cdots \epsilon 0]^T$

$R$  The column matrix whose elements are the measure numbers of a vector  $R$  from point  $O$ , the mass center of the spacecraft in its unstrained state, to  $Q$ , a point fixed in the rigid part of the spacecraft

$r$  The matrix  $[r^1 0 r^2 0 \cdots r^n 0]$ , where  $r^s$  is the  $3 \times 1$  matrix whose elements are the measure numbers of a vector  $r^s$  connecting point  $Q$  to a point  $Q_s$ , which is the mass center of the  $s$ th sub-body

$\sim$  This symbol is associated with a  $3 \times m$  matrix  $[a^1 a^2 \cdots a^m]$ , where  $a^i$  is a  $3 \times 1$  matrix. The tilde denotes the  $3m \times 3m$  matrix

$$\begin{bmatrix} \tilde{a}^1 & & 0 \\ & \tilde{a}^2 & \\ 0 & & \ddots \\ & & & \tilde{a}^m \end{bmatrix}$$

where  $\tilde{a}^i$  means the skew symmetric matrix

$$\begin{bmatrix} 0 & -a_3^i & a_2^i \\ a_3^i & 0 & -a_1^i \\ -a_2^i & a_1^i & 0 \end{bmatrix}$$

- $\dot{\theta}$  A column matrix with elements  $\dot{\theta}_1, \dot{\theta}_2, \dot{\theta}_3$  that are linear approximations to the measure numbers of the angular velocity  $\omega$  of the rigid base to which the appendages are attached
- $T$  Applied torque matrix
- $I$  The inertia matrix for the entire vehicle in its undeformed state
- $J$  A  $3 \times 3$  diagonal matrix whose elements are the polar moments of inertia of the three momentum wheels
- $\Omega^0$  The  $3 \times 1$  matrix whose elements are the nominal spin rates of the momentum wheels
- $\dot{\chi}$  A  $3 \times 1$  matrix whose elements are the momentum wheels' departures from  $\Omega^0$
- $\tau$  The  $3 \times 1$  matrix of torque applied to the polar moment of inertia axes of the momentum wheels

An equation which will be used in the sequel is that which is obtained by combining Eqs. (2) and (3):

$$T - \tau = (I - J) \ddot{\theta} - J \Omega^0 \dot{\theta} - \bar{\delta}^T \ddot{\eta} \quad (5)$$

### 3. Control Simulation

The Laplace transforms of Eqs. (1) and (5) are

$$[s^2 E + 2s \bar{\zeta} \bar{\sigma} + \bar{\sigma}^2] \bar{\eta}(s) = s^2 \bar{\delta} \theta(s) \quad (6)$$

$$T(s) - \tau(s) = [s^2 (I - J) - s \tilde{J} \Omega^0] \theta(s) - s^2 \bar{\delta}^T \bar{\eta}(s) \quad (7)$$

and substitution of the expression for  $\bar{\eta}(s)$  from Eq. (6) into Eq. (7) yields

$$T(s) - \tau(s) = [s^2 (I - J) - s \tilde{J} \Omega^0 - s^4 \bar{\delta}^T \bar{D} \bar{\delta}] \theta(s) \quad (8)$$

where

$$\bar{D} = (s^2 E + 2s \bar{\zeta} \bar{\sigma} + \bar{\sigma}^2)^{-1} = \begin{bmatrix} \frac{1}{s^2 + 2s \zeta_1 \sigma_1 + \sigma_1^2} & & \\ & \frac{1}{s^2 + 2s \zeta_2 \sigma_2 + \sigma_2^2} & \\ & & \ddots \\ & & & \frac{1}{s^2 + 2s \zeta_N \sigma_N + \sigma_N^2} \end{bmatrix} \quad (9)$$

The symbol  $N$  denotes the number of modal coordinates  $\bar{\eta}_i$  retained after truncation.

If  $G(s)$  is defined by

$$G(s) = [s^2 (I - J) - s^4 \bar{\delta}^T \bar{D} \bar{\delta} - s \tilde{J} \Omega^0]^{-1} \quad (10)$$

then,

$$\theta(s) = G(s) [T(s) - \tau(s)] \quad (11)$$

and  $G(s)$  is referred to as the matrix of transfer functions.

Figure 3 depicts the vehicle control system block diagram while Fig. 4 shows the details of the dynamics block. It is apparent from Fig. 4 that

$$T(s) - \tau(s) = [R(s)^{-1} - F(s)] \theta(s) \quad (12)$$

Hence, from Eqs. (12) and (8),

$$R(s) = [s^2 (I - J) - s \tilde{J} \Omega^0]^{-1} \quad (13)$$

$$F(s) = s^4 \bar{\delta}^T \bar{D} \bar{\delta} \quad (14)$$

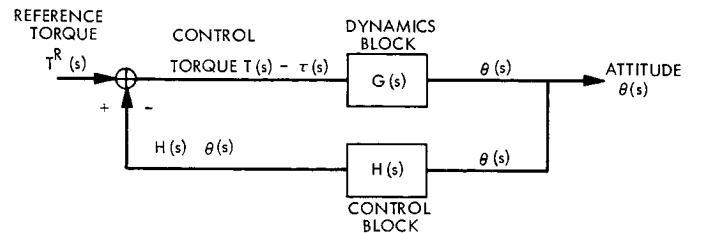


Fig. 3. Vehicle control system

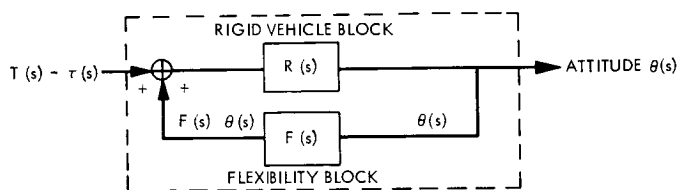


Fig. 4. Dynamics block

An alternative form for Eq. (14) is

$$F(s) = s^4 \sum_{i=1}^N \frac{\delta_i^T \delta_i}{s^2 + 2s\zeta_i\sigma_i + \sigma_i^2} \quad (15)$$

Then, in more detail,  $F(s)$  can be depicted as shown in Fig. 5.

For the case when the Canopus tracker is mounted on the rigid spacecraft bus, the relationship between the reference torque,  $T^R(s)$ , and  $\theta(s)$  is

$$T^R(s) = [G^{-1}(s) + H(s)] \theta(s) \quad (16)$$

In the event that the tracker is attached to a flexible appendage, it will sense the rotations  $\beta^i$  of the appendage's  $i$ th sub-body, to which it is attached, as well as  $\theta$ , the rotations of the rigid component of the spacecraft. Modification of Eq. (16) to account for  $\beta^i$  is accomplished as follows. Recall that

$$q = [u_1^1 u_2^1 u_3^1 \beta_1^1 \beta_2^1 \beta_3^1 \cdots u_1^i u_2^i u_3^i \beta_1^i \beta_2^i \beta_3^i \cdots u_1^n u_2^n u_3^n \beta_1^n \beta_2^n \beta_3^n]^T = \bar{\phi} \bar{\eta} \quad (17)$$

$\beta^i$  is the  $(2i)$ th submatrix of dimension  $3 \times 1$  in  $q$ . If the  $(2i)$ th submatrix of dimension  $3 \times N$  in  $\bar{\phi}$  is denoted  $\bar{\phi}_{2i}$ , then

$$\beta^i = \bar{\phi}_{2i} \bar{\eta} \quad (18)$$

Instead of measuring  $\theta$ , the Canopus tracker measures  $\theta + \bar{\phi}_{2i} \bar{\eta}$ . But,

$$\theta(s) + \bar{\phi}_{2i} \bar{\eta}(s) = [E + \bar{\phi}_{2i} s^2 (s^2 E + 2s\bar{\zeta}\bar{\sigma} + \bar{\sigma}^2)^{-1} \bar{\delta}] \theta(s) \quad (19)$$

Hence, the modified form of Eq. (16) is

$$T^R(s) = \{G^{-1}(s) + H(s) [E + \bar{\phi}_{2i} s^2 (s^2 E + 2s\bar{\zeta}\bar{\sigma} + \bar{\sigma}^2)^{-1} \bar{\delta}]\} \theta(s) \quad (20)$$

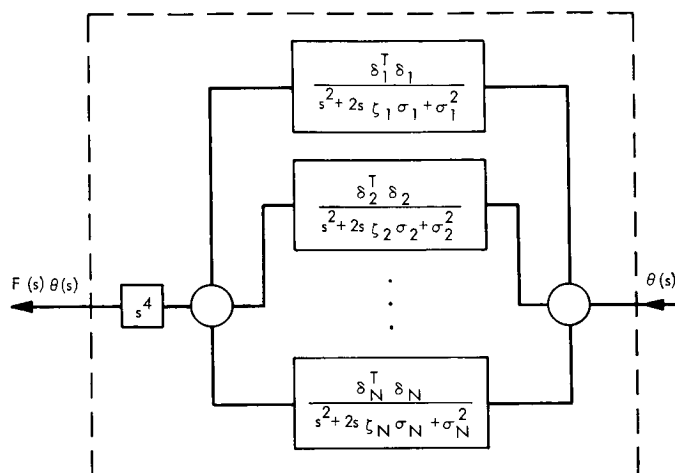


Fig. 5. Flexibility block

Equation (20) must be used in place of Eq. (16) for the case when a sensor is attached to a flexible appendage.

#### 4. Calculations for the TOPS Configuration

The design shown in Fig. 2 has been subjected to structural analysis, by means of the theory of finite elements, by R. Bamford and C. Helberg of the Engineering Mechanics Division. Their model has a total of 351 degrees of freedom. Thus, a maximum of 351 mode shapes can be derived. However, after truncations, only the first 30 modes were retained. The SAMIS<sup>1</sup> structural dynamics program was used to derive the data required by Eq. (20), that is, the mode shapes  $\bar{\phi}$ , the natural frequencies  $\bar{\sigma}$ , and the rigid elastic coupling terms  $\bar{\delta}$ .

Rather than undertake a lengthy parametric study of Eq. (19), which would require knowledge of the control law, it is possible to obtain a general idea of the influence of flexibility on pointing accuracy by examining the first mode response to a sizeable deflection of that part of the vehicle experiencing the largest deflection in that mode. That is, the assumption is made that

$$\bar{\eta} = [\eta_1 0 0 \cdots 0]^T \quad (21)$$

The node experiencing largest deflection for the condition in Eq. (21) is determined by the largest number in the first column of  $\phi$  (i.e., the largest number in the first eigenvector). That number for the TOPS vehicle is 0.4535, and corresponds to a deflection of node 306 in the  $x_3$  direction, i.e.,

$$u_3^{306} = 0.4535 \eta_1 \quad (22)$$

<sup>1</sup>SAMIS = Structural Analysis and Matrix Interpretive System.

Node 306 is located at the end of the long flexible boom supporting the magnetometer, and, for this reason, it could have been guessed that this node would experience large deflection. Selection of a value for  $u_3^{306}$  uniquely determines values for  $\beta_i^{69}$ , the rotations at node 69. Node 69 is located at the tip of the truss structure where the Canopus tracker is situated. Thus,  $\beta_i^{69}$  are the rotations the Canopus tracker receives under the influence of the disturbance producing  $u_3^{306}$ . More explicitly, from Eq. (18),

$$\beta_i^{69} = \phi_{2(69), i} \bar{\eta} \quad (23)$$

and from Eq. (22)

$$\beta_i^{69} = \phi_{2(69), i} \frac{u_3^{306}}{0.4535} \quad (24)$$

Values for  $\phi_{2(69), i}$  are available from data provided by the structural analysis. Using this information

$$\beta_1^{69} = -3.85 \times 10^{-7} u_3^{306} \quad (25)$$

$$\beta_2^{69} = 3.81 \times 10^{-8} u_3^{306} \quad (26)$$

$$\beta_3^{69} = 5.64 \times 10^{-9} u_3^{306} \quad (27)$$

For a large deflection  $u_3^{306}$ , say 10 inches,  $\beta_i^{69}$ ,  $i = 1, 2, 3$ , is

$$\beta_1^{69} = -2.20 \times 10^{-4} \text{ deg} \quad (28)$$

$$\beta_2^{69} = 2.18 \times 10^{-5} \text{ deg} \quad (29)$$

$$\beta_3^{69} = 3.23 \times 10^{-6} \text{ deg} \quad (30)$$

With such small rotations at node 69, it is felt that flexibility will have extremely small influence on sensing accuracy, and the idea of placing the Canopus tracker at the end of the truss structure is feasible.

## Reference

1. Likins, P. W., *Dynamics and Control of Flexible Space Vehicles*, Technical Report 32-1329, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1969.

## C. Failure Detection System for Solenoid-Actuated Gas Jet Valves, F. G. Roselli-Lorenzini

### 1. Introduction

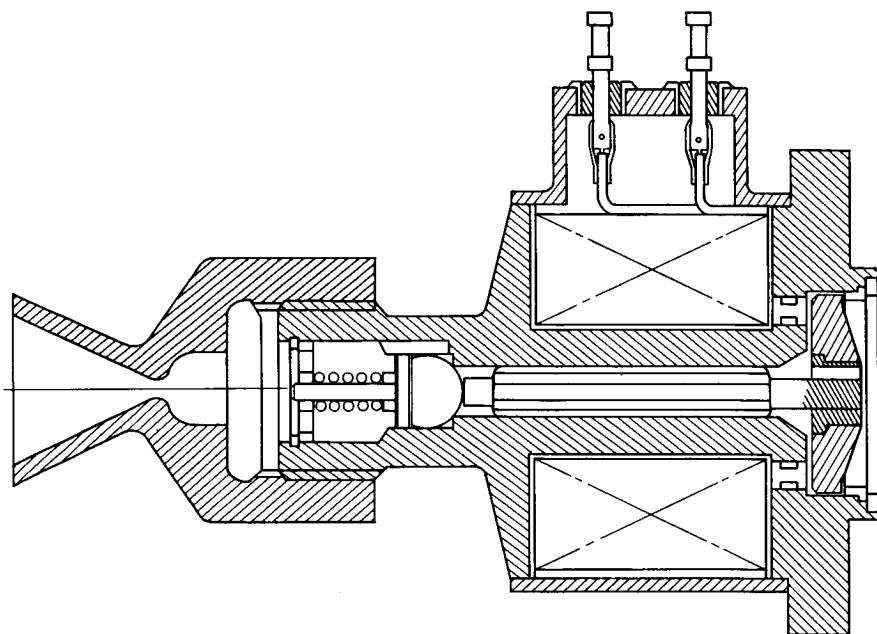
The thrust-supplying elements of the *Mariner*-type cold gas reaction control system are 12 solenoid-actuated jet valves. A nozzle is applied at the outlet of each valve. The complete thruster is shown in Fig. 6. The thrusters

are operated in pairs, providing the spacecraft with pure couples about the center of gravity. They perform all the spacecraft attitude-control torquing functions, including injection rate reductions, celestial reference acquisitions, commanded turns, limit cycle cruise attitude control, and roll control during midcourse vector burns. The active life of the spacecraft, in absence of failures, is determined by the active life of the reaction control system, i.e., the amount of propellant the system stores at launch and its rate of consumption. As soon as the propellant vessels are depleted, the spacecraft is no longer in control and starts to tumble in space, thus reducing the efficiency of the communication links to the point where the spacecraft is considered lost. Any loss of propellant after launch is reflected as a reduction of the spacecraft active life; a component failure leading to a significant dumping of gas in space is considered to be critical. Potential failures of this nature are almost exclusively associated with the jet valves. A perfectly sealing jet valve seat is a technical improbability: a certain amount of leakage is always allowed, quantitatively defined by the characteristics of the seat. Thus, a failure mode is a leak rate out of specification. The only other relevant failure mode is a valve stuck open after an actuation.

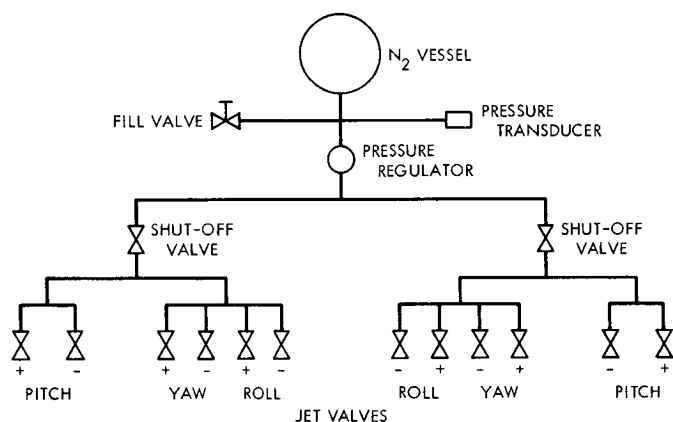
The *Mariner*-type reaction control system includes a built-in redundancy to reduce the risk of such failure modes. Two propellant vessels are provided, each serving half of the thrusters; each half-system is able to provide a complete three-axis control of the spacecraft, although supplying only torques instead of pure couples. Moreover, each vessel carries enough propellant to complete the mission, given a failure at launch, and to compensate for the torque supplied by an open-valve failure of the other half-system. The weight performance of such a "dual-redundant" system is, of course, very poor. An alternate, more weight-effective configuration is shown in Fig. 7. The two half-systems are now linked to the same pressure vessel. As soon as an open-valve failure occurs in one of the half-systems, an initially open valve is activated to shut off the failed half-system feed line, leaving the spacecraft in control of the other half-system only. This configuration allows a considerable weight savings for both propellant and hardware, but its feasibility depends on the availability of a failure detection system to control the in-line isolation valve. A promising approach in designing such a failure detection system is described below.

### 2. Functional Description and Test Results

A series of advanced development programs performed by the General Electric Co., Space Systems Division,



**Fig. 6. Mariner IV thruster**

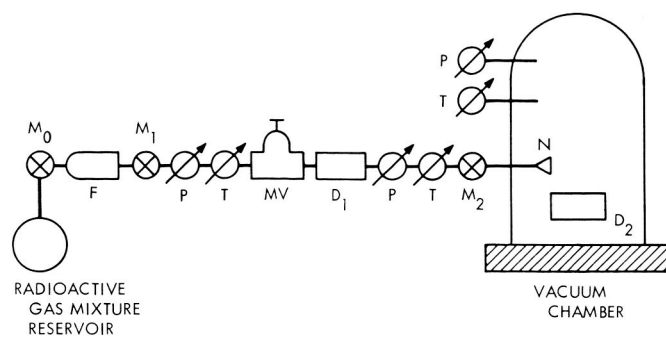


**Fig. 7. Reaction control system alternate configuration**

Valley Forge, Pa., in the past few years led to the production of an avalanche-type radiation detector. The detector consists basically of an avalanche diode and a tunnel diode amplifier. When radiation is absorbed in a conventional silicon diode, it creates hole-electron pairs which through their own mobility are swept across the *PN* junction of the silicon diode, resulting in a current flow. The avalanche diode has a high internal field, and each time the radiation produces an electron pair, they in turn ionize silicon atoms. If the internal field is high enough, there is a cascade of electrons from each absorption. The tunnel diode coupled to the avalanche diode triggers each time the junction current totals 400 electrons. The result is a digital pulse for each radiation absorption, making

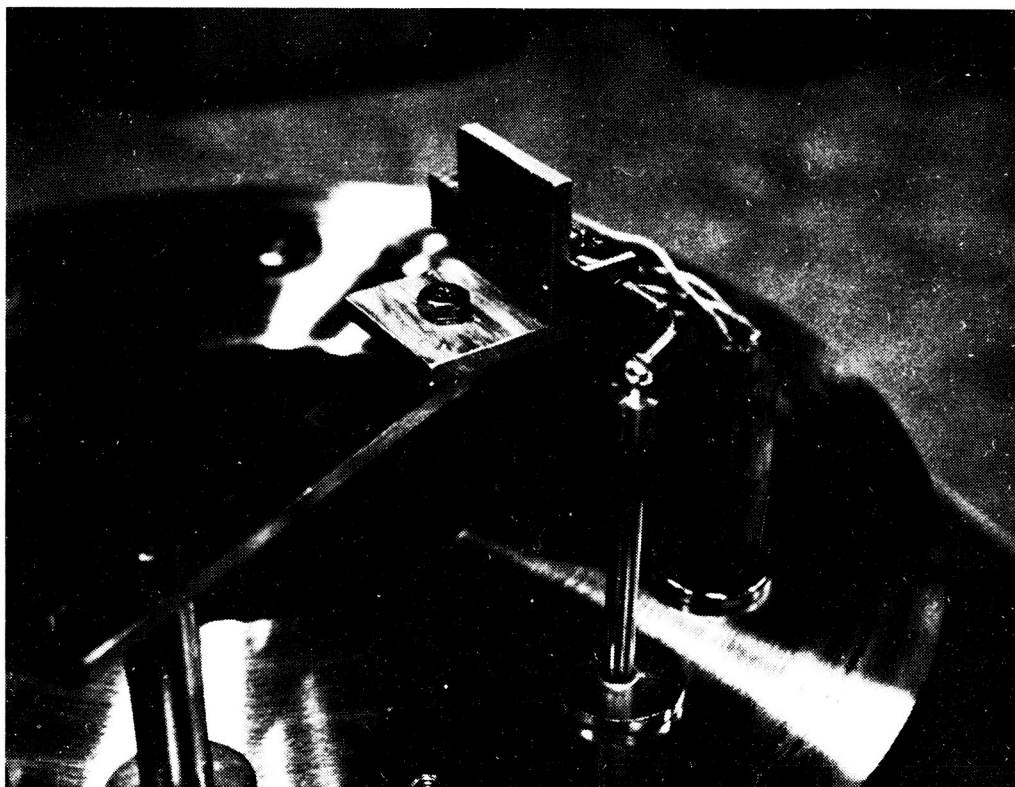
this couple the solid-state analog of the Geiger counter. A radiotracer, such as krypton (Kr 85) or carbon 14 (C14), may be added in small quantities to the cold gas propellant and an avalanche detector appropriately placed and calibrated to monitor the jet valve leak rate of the radiotracer "as viewed" by the avalanche diode. If the number of counts in a prefixed length of time exceeds a given set threshold value, an on-board closed-loop system could actuate a valve and cut off the half-system to which the failed valve belongs.

A Phase I contract granted to the General Electric Co. in 1969 assessed the feasibility of the concept. Initially aimed to ascertain the performance of an avalanche detector system placed outside the nozzle in view of the plume, it has evolved to a detector system "looking" into the nozzle feed line. In fact, it was soon evident that the detection capability of the outside detector was too low to discriminate against the background noise level counting. The pneumatic test apparatus outline is shown in Fig. 8; the outside detector mounted in a vacuum chamber is shown in Fig. 9a. The radiotracer was C14, labeled carbon monoxide (CO), in a 10% CO mixture with nitrogen. The detector was calibrated using an MV-25-XL series VACCO metering valve, having a calibration range from  $10^{-10}$  to  $5 \text{ cm}^3 (\text{std})/\text{s}$ , as the leaking element. Upstream and downstream pressures and temperatures were monitored. The calibration was performed by tabulating the number of C14 emissions ( $C_1$ ) detected in counts/min as a function of the gas flow rate ( $F$ ) in  $\text{cm}^3/\text{s}$ . The data



$M_1$     MANUAL VALVES  
 F       FILTER  
 P       PRESSURE TRANSDUCER  
 T       TEMPERATURE TRANSDUCER  
 MV      METERING VALVE  
 N       NOZZLE  
 $D_1, D_2$     AVALANCHE DETECTORS

**Fig. 8. Pneumatic test apparatus**



**Fig. 9. Outside detector mounted in vacuum chamber**

is then normalized to the sensitive volume ( $V_s$ ), i.e., the volume of gas the detector views in the tubing, the activity of the radioactive gas ( $A_i$ ), and the pressure. The calibration factor takes the form

$$\frac{C_1}{V_s} \left( \frac{A_i}{F} \times \frac{P_t}{P_i} \right)^{-1} = \frac{\text{counts}}{\mu\text{Ci}} : \frac{\text{cm}^3(\text{std})}{\text{h}}$$

where  $P_t$  is the pressure in the test volume, and  $P_i$  is the pressure in the gas reservoir. The results obtained are as follows:

Leak rate, $\text{cm}^3(\text{std})/\text{h}$	Detector calibration factor	Total error, %
0.72	$4.87 \times 10^4$	$\pm 50$
1.44	$1.25 \times 10^4$	$\pm 35$
4.32	$0.72 \times 10^4$	$\pm 28$

The total error is cumulative of the errors in the counting statistics, the pressure readings, the C14 radioactivity standard, and the metering valve calibration procedure.

### 3. Conclusions

The test results have shown the feasibility of the concept for leak detection in cold gas jet valves. A considerable improvement in accuracy is possible by both refining the test technique, thus eliminating potential sources of errors, and increasing the activity of gas and/or the counting time. Phase II of the program has been started to generate a breadboard, flight-oriented leak detector included in the thruster nozzle and viewing the volume between the valve seat and the nozzle throat section (Fig. 10).

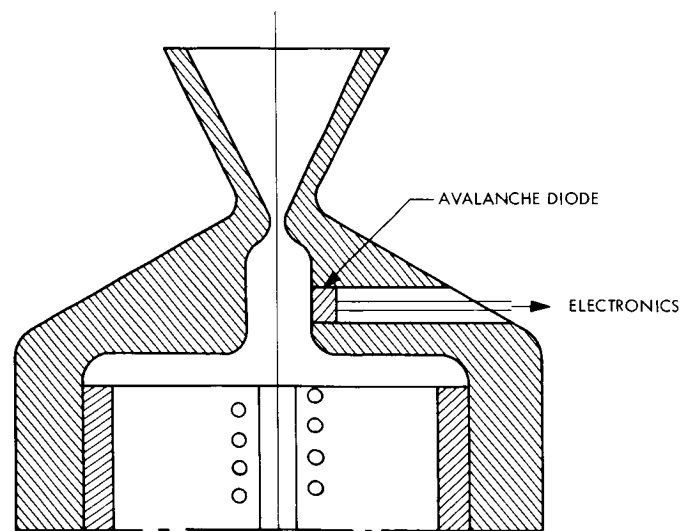


Fig. 10. Proposed flight-borne leak detector location

A considerably increased amount of testing over Phase I is planned to allow a more accurate statistical evaluation of the test results. A Phase III is also planned to build a breadboard of the closed-loop failure detection and recovery system. In such a system, the leak detector monitors the leak rate of the valve at discrete time intervals, being disabled when the valve is operated. As the valve fails and its leak rate increases over a specified threshold, a signal from the detector actuates a shut-off valve in the line. The system will be proposed for consideration in designing new reaction control systems, such as for the Viking 1975 orbiter-class vehicles.

## D. Wiener-Type Filtering for Multivariable Nonlinear Processes, A. K. Bejczy

### 1. Introduction

Because of hardware considerations, the applicability of modern nonlinear filter theories to practical on-board guidance and control problems depends strongly upon the structural complexity of the filters. Constructing nonlinear filters having "manageable" structures is, therefore, an important research and development goal.

In two previous articles (SPS 37-59, Vol. III, pp. 132-137, and SPS 37-60, Vol. III, pp. 112-118), a method was briefly described and illustrated for evaluating the performance of approximate nonlinear filters. It was also pointed out and demonstrated that the new performance evaluation method can also be conveniently extended to construct simple, approximate nonlinear filter schemes. Conceptually, the new performance evaluation and filter construction method for nonlinear processes is based upon assumed steady-state operational behavior (steady-state error statistics), which is one of the distinguishing features of the Wiener filter. The main points in the new method are (1) the employment of *constant filter gains*, and (2) the possibility of approximately evaluating the filters' performance by *deterministic techniques*. Furthermore, (3) suitable constant filter gains can be determined, aided by the (Wiener-type) requirement of *minimum steady-state variance* for the filtering error. Technically, however, the new filter construction and performance evaluation method, being based on differential equation techniques, follows modern filtering procedures.

In the latter reference (SPS 37-60, Vol. III), the new filter construction and performance evaluation method was applied to a two-dimensional nonlinear problem from the point of view of control system design analysis. In the

present article the performance of the simple, minimum steady-state variance constant gain (MSSVCG) nonlinear filters will be compared to the performance of the much more complex, full, maximum principle least squares (MPLS) nonlinear filters. The aim of this comparative study is to provide deeper insight into the ramifications of the MSSVCG nonlinear filters.

## 2. MPLS Nonlinear Filter for a Forced Nonlinear Spring System

In this article the spring system, the measurement vector, and the system and measurement disturbances, as well as the notations, will be the same as those considered and described in *Subsection 4*, SPS 37-60, Vol. III, p. 113, where the pertinent equations are Eqs. (1-6).

Applying the MPLS nonlinear filter theory (Ref. 1) to the problem under consideration, the following filter equations are obtained:

$$\dot{\hat{x}}_1 = \hat{x}_2 + P_{11} Q_{11}^{-1} [y_1 - \hat{x}_1] + P_{12} Q_{22}^{-1} [y_2 - \hat{x}_2] \quad (1)$$

$$\begin{aligned} \dot{\hat{x}}_2 = & -2\hat{x}_1 - a\hat{x}_1^3 - 3\hat{x}_2 + 5 \sin(t) \\ & + P_{12} Q_{11}^{-1} [y_1 - \hat{x}_1] + P_{22} Q_{22}^{-1} [y_2 - \hat{x}_2] \end{aligned} \quad (2)$$

$$\dot{P}_{11} = 2P_{12} - P_{11}^2 Q_{11}^{-1} - P_{22}^2 Q_{22}^{-1} + R_{11} \quad (3)$$

$$\begin{aligned} \dot{P}_{12} = & P_{22} - 3P_{12} - (2 + 3a\hat{x}_1^2) P_{11} - P_{11} P_{12} Q_{11}^{-1} \\ & - P_{12} P_{22} Q_{22}^{-1} \end{aligned} \quad (4)$$

$$\begin{aligned} \dot{P}_{22} = & -6P_{22} - (4 + 6a\hat{x}_1^2) P_{12} - P_{12}^2 Q_{11}^{-1} \\ & - P_{22}^2 Q_{22}^{-1} + R_{22} \end{aligned} \quad (5)$$

where  $a = 0.5$  is used in the calculations, and the initial conditions are given as

$$\hat{x}_1(0) = a$$

$$\hat{x}_2(0) = a_2$$

$$P_{11}(0) = b_{11}$$

$$P_{12}(0) = b_{12}$$

$$P_{22}(0) = b_{22}$$

In the above equations, the hat ( $\hat{\cdot}$ ) signifies the estimated (mean) value of the state variables  $x_i$ , the dot ( $\dot{\cdot}$ ) denotes time derivative, and  $P_{ij}$  is the approximate covariance of the filtering errors  $x_i - \hat{x}_i$ . In Eqs. (1) and (2), the  $P_{ij} Q_{ii}^{-1}$  terms are the filter gains operating on the differences  $y_i - \hat{x}_i$ .

From the point of view of mechanization and performance evaluation of the filter, the significant feature of Eqs. (1-5) is the fact that they are *coupled, stochastic* differential equations. Thus, the filter gains can not be predetermined; they are computed on-line, that is, their computation is a part of the running estimation procedure. Furthermore, the performance of the filter (that is, the approximate covariance of the filtering error) must be evaluated by Monte Carlo techniques since now  $P_{ij}$  itself is a stochastic quantity.

It is worthwhile to note that, for the problem under consideration, the MPLS filter equations (Eqs. 1-5) are formally equivalent to using linear filtering about the computed mean, and they are also equivalent to the minimum variance filter equations for a linear expansion of the system's nonlinearities. This is so because of the linear character of the measurement vector that is considered in the present problem.

## 3. MSSVCG Nonlinear Filter for a Forced Nonlinear Spring System

For the problem under consideration, the relevant MSSVCG filter equations are given in a previous study (see Eqs. 7-15, SPS 37-60, Vol. III, p. 114). There, the filter gains are given by Eq. (12), computed from Eqs. (13-15); the performance of the filter (that is, the approximate covariance of the filtering error) is computed from Eqs. (9-11), while the filter equations for estimating the mean of the state variables  $x_i$  are given by Eqs. (7) and (8). Equations (7-11) of SPS 37-60, Vol. III, p. 114, should be compared to Eqs. (1-5) of this article.

As seen and emphasized in SPS 37-60, Vol. III, the significant features of the MSSVCG filter equations are as follows:

- (1) The filter gain is precomputed; it is a constant (deterministic) quantity. (Note that in the MPLS filter scheme the filter gain itself is a stochastic quantity.)
- (2) The performance of the filter is described approximately by deterministic equations.

- (3) The mechanization of the MSSVCG filter requires the implementation of only two equations. (Note that the mechanization of the corresponding MPLS filter requires the implementation of five equations.)

It must be emphasized that, despite the Wiener-type requirement involved in the conceptual framework of constructing the MSSVCG nonlinear filter, the technical evaluation of the gain and/or performance of the MSSVCG nonlinear filter does not require the manipulation with Wiener-type integral equations. The mathematical operations involved in determining the MSSVCG nonlinear filter gain and/or performance are (1) solving a system of quadratic algebraic (or, alternatively, differential) equations, and (2) solving a system of linear differential equations. Thus, the technical features of the MSSVCG nonlinear filter scheme are, in general, much more convenient than the technical features of a Wiener-type integral equation.

#### 4. Numerical Results

As pointed out above, the gain and/or performance of the MPLS filter must be evaluated by Monte Carlo techniques. [The Monte Carlo digital simulation procedure for a sequential filter is briefly described in a previous article (SPS 37-59, Vol. III). The only change now is the increase of the number of variables.] All Monte Carlo simulation results quoted below are calculated from 200 samples, that is, from 200 realizations of the filter. In the simulation procedure the sampling time was made equal to the integration stepsize  $\Delta t$ . ( $\Delta t = 0.1$  s was used.) On the computer plots and the diagrams of this article, however, only points with 0.5-s intervals are shown using linear interpolation between them. Since the filter, Eqs. (1-5), is formulated in the continuous time domain, the continuous Gaussian random processes specified by  $Q_{ii}$  and  $R_{ii}$  had to be translated into equivalent discrete random sequences  $\bar{Q}_{ii}$  and  $\bar{R}_{ii}$  in the digital simulation procedure. This is made through the equivalence claim

$$\bar{Q}_{ii} = \frac{Q_{ii}}{\Delta t}$$

and

$$\bar{R}_{ii} = \frac{R_{ii}}{\Delta t}$$

In subsequent calculations it is assumed that  $R_{11} \equiv 0$ . That is, only the acceleration is disturbed by system noise. Furthermore, the MPLS and MSSVCG filters

are calculated and/or simulated for two measurement configurations:

- (1) There are only position measurements, in which case terms containing  $Q_{22}$  drop out from Eqs. (1-5).
- (2) There are both position and velocity measurements, in which case Eqs. (1-5) are fully applied.

**Case 1: Only position measurements.** For this case, the relevant equations for the MSSVCG filter are given in SPS 37-60, Vol. III:

- (1) Equation (16) yields the value of the constant filter gain for the minimum steady-state variance.
- (2) Equations (9-11), with  $G_{12} = G_{22} \equiv 0$ , yield the performance characteristics of the filter.
- (3) Equations (7) and (8), again with  $G_{12} = G_{22} \equiv 0$ , yield the estimate of the mean of the state variables  $x_i$ .

Figure 11 depicts the performance characteristics of the MPLS and MSSVCG filters in terms of their (approximate) covariance for  $Q_{11} = 1$  and  $R_{22} = 2$ . In this and all subsequent figures of this article,  $P_{ij}$  refers to the MPLS filter while  $\beta_{ij}$  refers to the MSSVCG filter. For comparison, both the "theoretical" and the "experimental" values of  $P_{ij}$  are depicted in this figure. The curves called " $P_{ij}$  theoretical" are actually the mean values of 200 numerical "solutions" of Eqs. (3-5). The curves called " $P_{ij}$  experimental" are computed from the 200 numerical "solutions" of Eqs. (1) and (2), using the standard formulas for the empirical covariance. The " $\beta_{ij}$  theoretical" curves are solutions of Eqs. (9-11) of SPS 37-60, Vol. III, with the relevant values of the minimum steady-state variance constant filter gains.

As seen in Fig. 11, the "theoretical" and "experimental" values of  $P_{ij}$  agree almost completely. There is also substantial agreement between the " $\beta_{ij}$  theoretical" and " $P_{ij}$  theoretical" curves in the following sense: the " $\beta_{ij}$  theoretical" curves are either nearly bounding or nearly centering the " $P_{ij}$  theoretical" curves. It must be observed that the time and/or state dependent "sub-structure" of the filter statistics is not reproduced by the " $\beta_{ij}$  theoretical" curves. This is to be expected, however, since the  $\beta_{ij}$  equations (Eqs. 9-11 of SPS 37-60, Vol. III) do not contain time and/or state dependent coefficients; the  $\beta_{ij}$  equations reflect only on the time and/or state independent characteristics of the nonlinear system in question.

Figure 12 compares the " $P_{ij}$  experimental" curves for the same case described above in connection with Fig. 11.

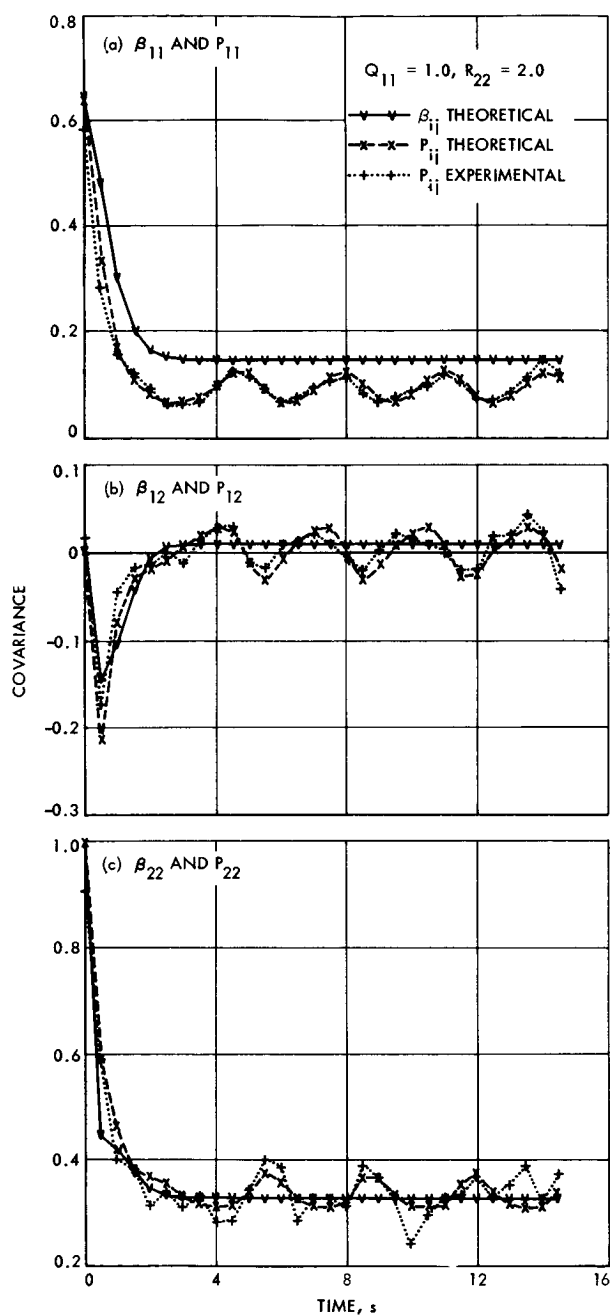


Fig. 11. Theoretical values of the MSSVCG and MPLS filters' performance (partial state observations)

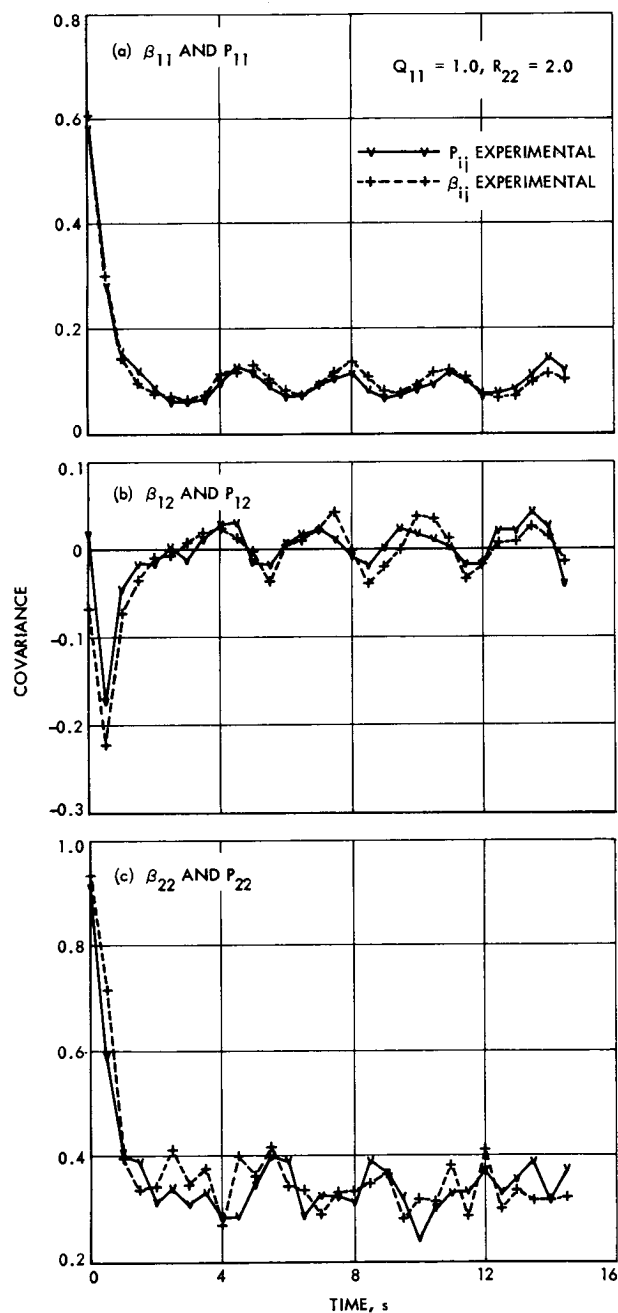


Fig. 12. Experimental values of the MSSVCG and MPLS filters' performance (partial state observations)

The " $\beta_{ij}$  experimental" curves are computed from the 200 numerical "solutions" of Eqs. (7) and (8) of SPS 37-60, Vol. III, using the appropriate values of the minimum steady-state variance constant filter gains. As seen in this figure, the experimental performance curves of the MPLS and MSSVCG filters for the problem under consideration are almost identical. Thus, for all practical purposes, the performance of the simple MSSVCG filter is as good as the much more complex MPLS filter scheme.

**Case 2: Position and velocity measurements.** For this case, the relevant equations for the MSSVCG filter are also given in SPS 37-60, Vol. III:

- (1) Equations (13-15) yield the values of the constant filter gains for the minimum steady-state variance.
- (2) Equations (9-11) yield the performance characteristics of the filter.
- (3) Equations (7) and (8) yield the estimate of the mean of the state variables  $x_i$ .

Figure 13 displays the performance characteristics of the MPLS and MSSVCG filters in terms of their (approximate) covariance for  $Q_{11} = Q_{22} = R_{22} = 1$ . Here again, it is seen that the "theoretical" and "experimental" values of  $P_{ij}$  agree almost completely. The " $\beta_{ij}$  theoretical" and " $P_{ij}$  theoretical" curves also exhibit a substantial agreement in the sense that the " $\beta_{ij}$  theoretical" curves are either closely bounding or closely centering the " $P_{ij}$  theoretical" curves.

In Fig. 14, the " $\beta_{ij}$  experimental" curves are compared to the " $P_{ij}$  experimental" curves for the case described above in connection with Fig. 13. Here, again, the experimental performance curves of the MPLS and MSSVCG filters for the problem under consideration are nearly identical. This means that, for all practical purposes, the performance of the simple MSSVCG filter is as good as the performance of the much more complex MPLS filter scheme. The fact that the " $\beta_{ij}$  theoretical" curves do not reproduce the "fine structure" of the filter statistics is to be expected since the  $\beta_{ij}$  equations take into account only the time and/or state independent characteristics of the nonlinear system under consideration.

Figure 15 displays one sample case of the MPLS filter and one sample case of the MSSVCG filter. These sample

cases are one of the 200 actual realizations of the two filter schemes by Monte Carlo simulation on a digital computer. This figure depicts the "true" disturbed velocity and position trajectories and the "estimated" velocity and position trajectories, together with the  $\sigma$  (standard deviation) channels centered around the estimated trajectories. The  $\sigma$  channels are computed from the solutions of Eqs. (9-11) of SPS 37-60, Vol. III, in the case of the MSSVCG filter and, in the case of the MPLS filter, they are computed from the mean of the solutions of Eqs. (3-5) of this article. Thus,

$$\sigma_{ii}(t) = [\beta_{ii}(t)]^{1/2}$$

or

$$\sigma_{ii}(t) = [P_{ii}(t)]^{1/2}$$

The noisy measurements are also shown in this figure.

Considering Fig. 15, the following observations can be noted:

- (1) Both the MSSVCG and MPLS filters estimate the mean of position and velocity very effectively despite the strong dynamical and measurement noise; they track the "true" trajectories after a short transient time and guarantee that reasonably large (85-90%) sections of the "true" trajectories are inside the filters' own  $\sigma$  channels, which are centered around the estimated mean values.
- (2) There is no noticeable difference between the two sample cases, one belonging to the MPLS and one to the MSSVCG filter. This is in perfect agreement with the overall performance characteristics of the two filters displayed and compared in Fig. 14. Thus, the comparison of the two sample cases also shows that the performance of the simple MSSVCG filter is as good as the performance of the more complex MPLS filter for the investigated multivariable nonlinear system.

#### Reference

1. Detchmendy, D. M., and Sridhar, R., "Sequential Estimation of States and Parameters in Noisy Nonlinear Dynamical Systems," *J. Basic Eng.*, Vol. 88, Series D, No. 2, pp. 362-369, 1966.

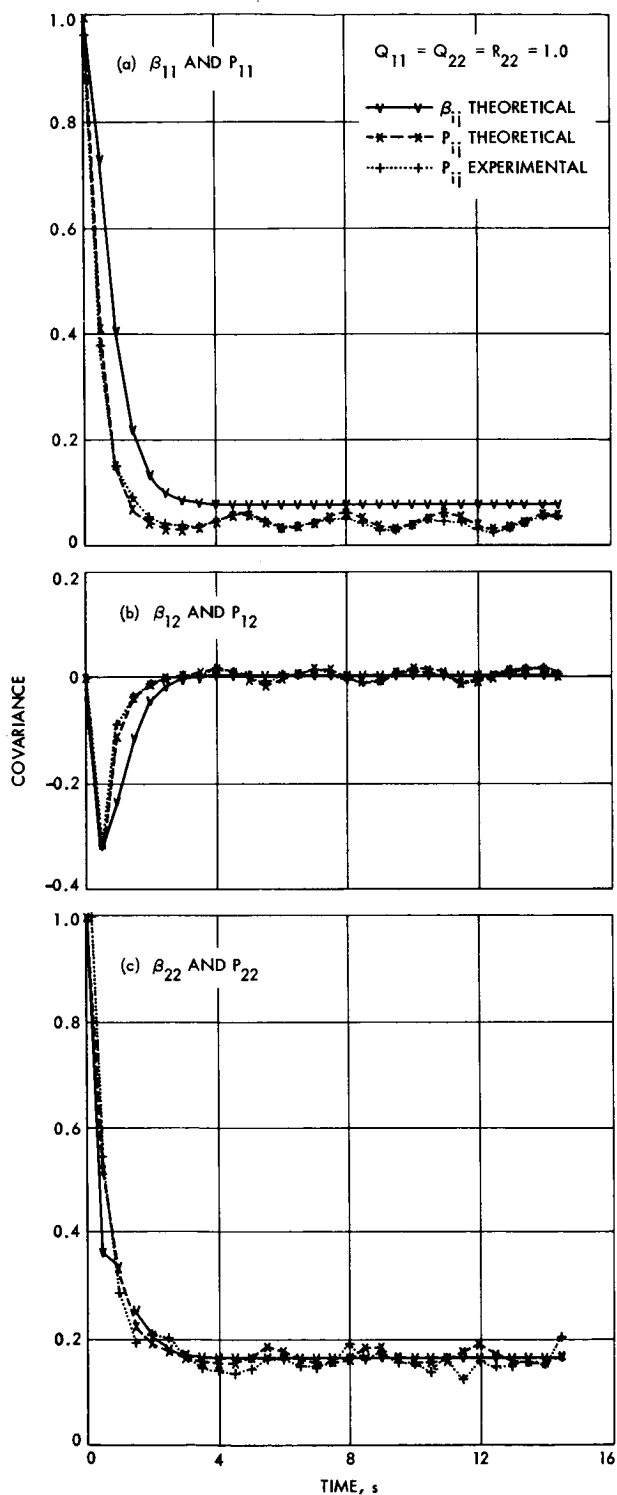


Fig. 13. Theoretical values of the MSSVCG and MPLS filters' performance (full state observations)

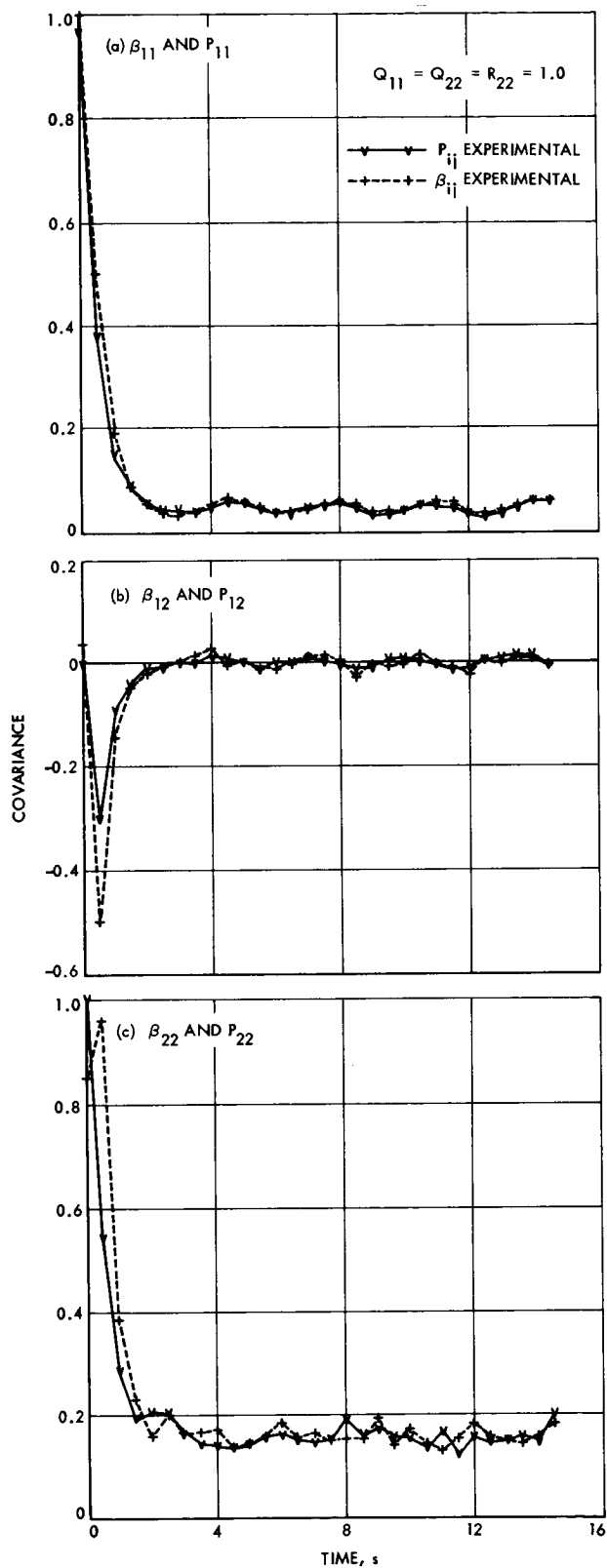


Fig. 14. Experimental values for the MSSVCG and MPLS filters' performance (full state observations)

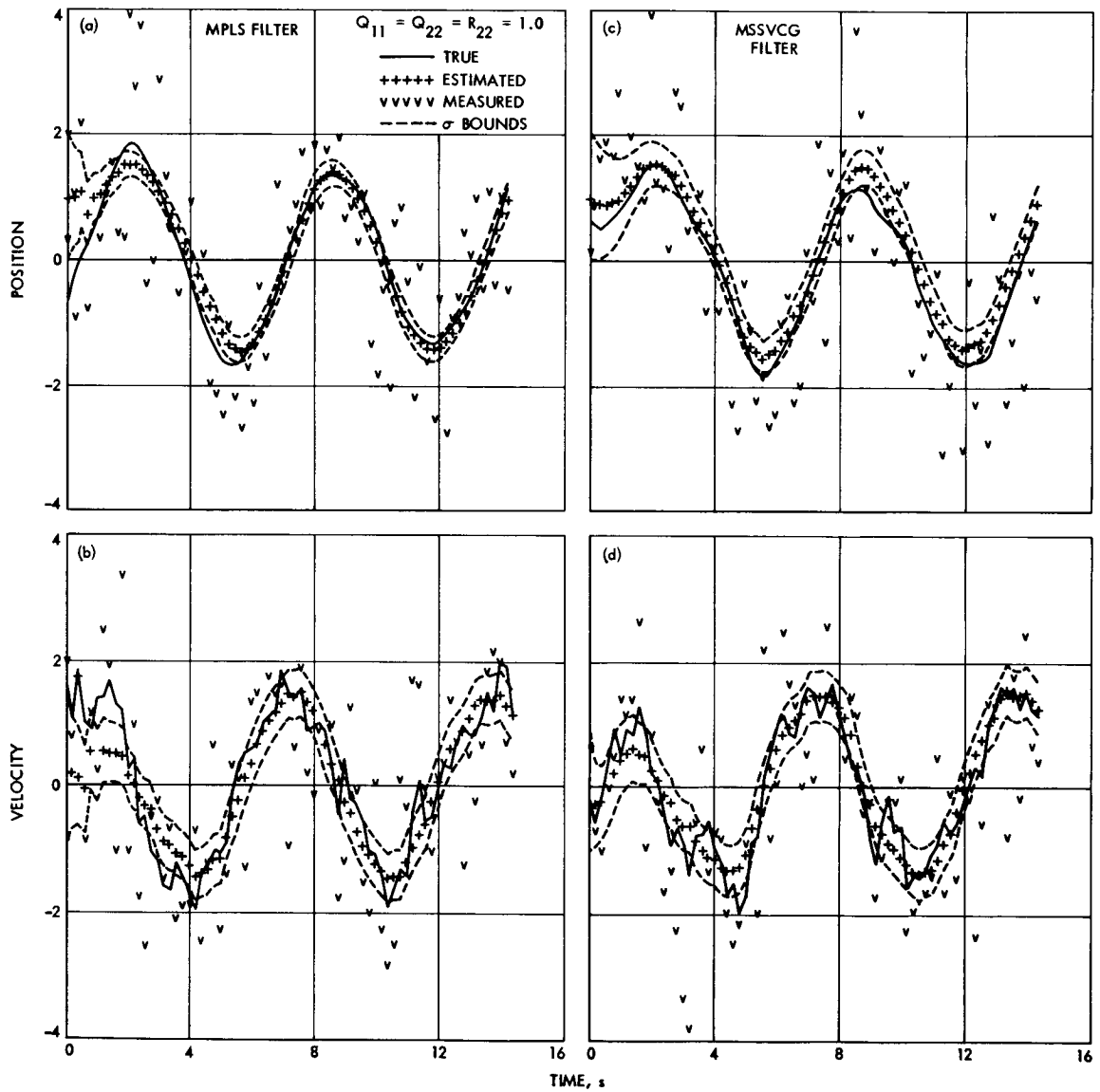


Fig. 15. Single Monte Carlo sample cases of the MSSVCG and MPLS filters' performance (full state observations)

## X. Materials

### ENGINEERING MECHANICS DIVISION

#### A. Survey of Microstrain Measuring Techniques for Use on Photoheliograph Mirror Material Evaluation, W. Rowe

##### 1. Introduction

Previous studies on the optical system for the *Apollo* Telescope Mount Photoheliograph have established a maximum acceptable deflection limit of  $10^{-7}$  in. for the primary mirror if the optics are to have near diffraction-limited performance capability. Such small deflection limitations make the mirror surfaces extremely sensitive to small strains. Some examples of these strains and their sources are: (1) elastic and plastic strains due to handling, the launch environment, and motional vibration; (2) thermal strains induced by the heat fluxes of the space environment; and (3) dimensional instability. Thus, a thorough understanding of the stress-strain behavior of the candidate primary mirror materials is a necessary prerequisite to the design and analysis of an effective space mirror system. To characterize these materials properly, the strain-measuring apparatus employed must be capable of reliably resolving deflections at least an order of magnitude smaller than the critical deflection ( $10^{-7}$  in.). Therefore, any method selected for use must have a strain sensitivity of at least  $10^{-8}$  in./in.

Although many investigators have performed microplasticity experiments in their studies of lattice frictional stresses, dislocation movements, and dimensional stability, their methods and results have varied. No standard technique has evolved which will consistently produce reliable microstrain data with a sensitivity of  $10^{-8}$  in./in.

##### 2. Purpose

The purpose of this portion of the work unit was to review and evaluate presently developed techniques on the basis of their ability to accurately and reliably measure the micro-level stress-strain behavior of candidate mirror materials. In addition, if required, the feasibility of modifying these techniques to obtain the necessary sensitivity level was to be determined. Based on the results, recommendations could be made with respect to the work to be performed.

The importance of knowing and measuring the material stability parameters is not overlooked. Knowledge of the effects of time and temperature on the dimensional stability of high acuity optics is also necessary for design purposes. The ability of the mirror surface to hold its optical figure is strongly dependent upon the dimensional stability of the mirror material. However, the scope of this

initial effort was purposely limited to exploring the possibility of investigating the stress-strain techniques. This will be accomplished initially by developing measuring techniques with sufficient sensitivity to obtain valid design data in the pre-macroyield region of the stress-strain curve. In addition it may also be necessary to know, very accurately, the magnitude of the stresses that are imposed on the specimens to account for those cases where the known elastic modulus data may be incomplete or invalid.

### 3. Technical Approach

The approach taken in performing this work unit was to first conduct a thorough review of known microstrain activities which was then followed by an evaluation of techniques applicable to the problem.

*a. Review of activities.* In order to perform the task of evaluating techniques, an extensive literature search was first performed on microplasticity and microstrain. This was followed by contacts to nearly 30 educational institutions, industrial laboratories, and government agencies that had been most active in the area of microstrain work. In this manner, information was obtained on the various methods employed and on possible modifications that might be needed to obtain the desired strain sensitivity of  $10^{-8}$  in./in. Visits were made to 16 of these organizations to evaluate their techniques and to determine their interests in the problem.

*b. Evaluation of techniques.* Following the extensive review and survey, an evaluation was made of the various techniques employed. This evaluation was based on information obtained during the review.

### 4. Results

The results of this survey and evaluation are discussed in the following paragraphs. In discussing the various methods employed, the finer details of the working principles of a particular technique are not given. This information is available in the literature and is beyond the scope of this report.

*a. Microstrain measurement techniques.* Microstrain measurements and experiments have been performed by many investigators using various techniques, and the results are well documented. Four basic methods have evolved: (1) the bonded resistance strain gauge, (2) the differential transformer, (3) the capacitance and (4) the electromechanical and mechanical-optical. Table 1 gives a broad comparison of these four methods based on their sensitivities, range, reversibility, and linearity. Since this

table was compiled, there have been refinements made to the differential transformer method by investigators at The Boeing Co. that have increased its apparent sensitivity in the torsion mode of loading to  $10^{-9}$  in./in.

For the purpose of this program, strain sensitivity is defined as the smallest measurable displacement that is detected in the specimen under test. It is dependent upon the gauge length of the specimen as well as the inherent sensitivity of the gauge. The range of a particular gauge means not only the difference between the maximum and minimum strain recorded, but also the capacity to intersperse the measurements of both large and small strains. The reversibility of a gauge is judged by how consistently the gauge can indicate zero deflection by the signal it generates, regardless of the direction from which the deflection is approached as long as the specimen under test behaves elastically. Linearity of response refers to a straight line relationship between the gauge deflection and the signal or response it generates to indicate displacement.

Besides the inherent sensitivity of the basic method employed, other considerations involved in gaining maximum strain sensitivity and reliability of measurements are minimizing background vibrations, drift of electrical equipment and thermal gradients. Some controversy and lack of agreement exist among the scientific community regarding the best method of loading the specimens; e.g., tension, compression, as well as the type of strain measuring device.

Table 1 shows that the capacitance and differential transformer methods appear to have the most potential, and during the survey portion of this effort they were found to be the most widely used microstrain measurement methods.

**Table 1. Comparison of various techniques for measuring microstrain**

Method	Sensitivity, in./in.	Range	Reversibility	Linearity
Electromechanical and mechanical-optical	$10^{-4}$ to $10^{-11}$	Variable	Poor	—
Resistance strain gauge	$0.5 \times 10^{-6}$	0 to 2%	Fair	—
Capacitance	$10^{-6}$ to $10^{-8}$	Unlimited	Excellent	Limited
Differential transformer	$10^{-5}$ to $10^{-6}$	Very wide	Excellent	Very good
This table developed by Dr. Norman Brown of the University of Pennsylvania.				

Another successful method has been used to evaluate the thermal and temporal stability of polished mirrors and mirror blanks, which is not shown in Table 1. This method involves the use of a Fizeau or other type of interferometer which can detect surface changes of a mirror to less than  $1 \times 10^{-8}$  in./in.

**b. Evaluation of techniques.** From the standpoint of sensitivity, the bonded resistance strain gauge with a maximum strain sensitivity of  $5 \times 10^{-7}$  in./in. does not meet the  $10^{-8}$  requirement for this program. Also, it does not allow reversibility of test runs and the cemented contacts introduce another questionable variable. An ordinary optical lever technique (Tuckerman optical gauge) also fails to provide the necessary sensitivity. The mechanical attachment required for the extensometer and limitations in the resolution of the usual manual reading of the mirror deflection are limiting factors. This method also has limited reversibility.

Apparent modification of the optical lever technique such as performed by Dr. Robert H. Chambers at the University of Arizona and Dr. Robert F. Tinder at Washington State University may provide sufficient sensitivities (approximately  $10^{-9}$ ), but the torsional method shear of loading does not provide data that can be readily correlated<sup>1</sup> to the type of stresses that the solar telescope optics will undergo.

The interferometric methods do not give the required stress-strain data but do provide valuable thermal and temporal stability information.

The linear variable differential transformer and the capacitance gauge were the most frequently used by the organizations contacted and visited. Seven, and possibly eight, of the organizations visited have the apparent potential to reach strain sensitivity levels of  $10^{-8}$  in./in. All but one of these organizations used the capacitance gauge technique. One organization has achieved the required  $10^{-8}$  in./in. sensitivity using torsional-type loading, but this could not be achieved by the desired tension or compression-type loading.

## 5. Conclusion

An evaluation of the results of this survey indicates that it should be feasible to modify some of the existing micro-strain measuring techniques to obtain the desired sensi-

<sup>1</sup>Theoretical methods for correlating tension, compression, and torsion data exist but have not been experimentally verified with high confidence.

tivity and reproducibility for micromechanical property determinations on candidate photoheliograph primary mirror materials.

## B. Long-Life Spacecraft Pressure Vessel Materials,

J. C. Lewis

### 1. Introduction

Establishing design allowable properties for materials for pressure vessels for long-life spacecraft requires the application of fracture mechanics principles to stress corrosive fluid-material combinations. The model being used where cyclic loading is negligible is

$$\left( \frac{K_{th}}{K_{Ic(pr)}} \right) \sigma_{pr} \geq \sigma_{ms} \quad (1)$$

where

$K_{th}$  = maximum applied stress intensity factor at which crack growth does not occur under sustained loading in the fluid-material-environment combination.

$K_{Ic(pr)}$  = fracture toughness of the material in the non-corrosive proof test environment.

$\sigma_{pr}$  = maximum stress achieved during the proof test.

$\sigma_{ms}$  = maximum allowable material stress at which crack propagation will not occur in the fluid-material-environment combination.

Since  $\sigma_{pr}$  is usually set to be the highest stress at which an uncracked pressure vessel will not fail in the proof-test environment, the testing is concentrated on the two fracture mechanics materials parameters,  $K_{th}$  and  $K_{Ic(pr)}$ . These parameters are determined for base metal, welds and heat-affected zones (HAZ) of the pressure vessel material.

### 2. Fluid-Material-Environment Combinations

The two fluid-material-environment combinations of interest for this program are nitrogen tetroxide (NTO)/monomethyl hydrazine (MMH) in Ti-6Al-4V alloy pressure vessels at 125°F and oxygen difluoride (OF<sub>2</sub>)/diborane (B<sub>2</sub>H<sub>6</sub>) at -180°F in some material yet to be determined. The NTO/MMH propellant combination is

typical of propulsion subsystems for such future missions as Venus-Mercury 1973, Viking 1975, Grand Tour, etc. The  $\text{OF}_2/\text{B}_2\text{H}_6$  propellant combination is typical of advanced propulsion subsystems for future orbiters of Jupiter and the other outer planets.

### 3. Test Procedures

The sustained load testing procedures developed by Tiffany and coworkers,<sup>2</sup> are being used with some modifications. Fatigue pre-cracked surface flawed specimens (Fig. 1) are being used for both  $K_{th}$  and  $K_{Ic(pr)}$ . Since proof testing Ti-6Al-4V alloy in liquid nitrogen ( $\text{LN}_2$ ) allows proof stresses above the ambient temperature strength of the material, determining  $K_{Ic(pr)}$  in  $\text{LN}_2$  on each NTO/MMH sustained load ( $K_{th}$ ) specimen is planned. Figure 2 is a schematic representation of the experimental setup. The selection of a proof-test environment for materials for  $\text{OF}_2/\text{B}_2\text{H}_6$  pressure vessels must await selection of specific materials. However, an effort will be made to obtain  $K_{Ic(pr)}$  values for each individual  $K_{th}$  specimen.

### 4. Equipment and Test Apparatus

Fatigue pre-cracking from an electrical discharge machined (EDM) notch is done on a 10,000-lb capacity, axially loaded fatigue machine which was acquired from government surplus, installed at JPL, and calibrated. The machine is stopped periodically to observe the crack

<sup>2</sup>Masters, J. N., Haese, W. P., and Bixler, W. D., "Fracture and Nitrogen Tetroxide/Sustained Load Flaw Growth of 6Al-4V Titanium," October 1968 (JPL internal document).

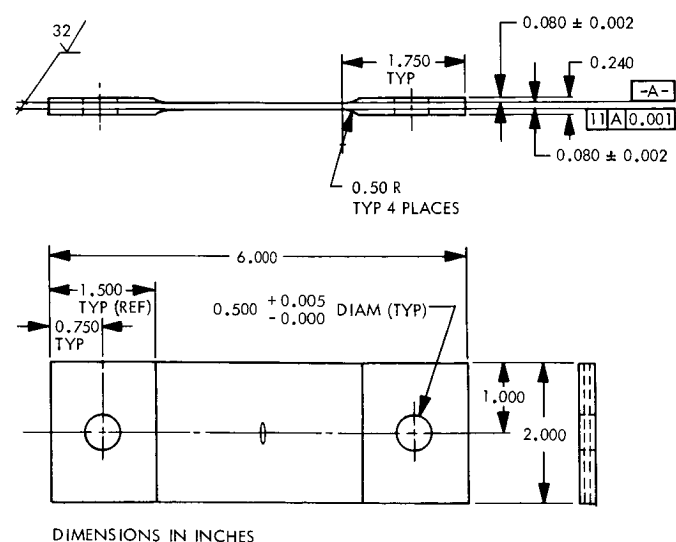


Fig. 1. Fracture mechanics test coupon

growth through a 40× telescope. Fatigue cracking is done to obtain a crack tip radius as sharp as a natural crack.

Four commercially available 20,000-lb sustained load-testing machines have been installed at the Edwards Test Station. Plumbing for NTO or MMH has been built to allow simultaneous testing of six tandem specimens per machine. An O-ring sealed cup maintains contact between the propellant and the crack zone, as shown schematically in Fig. 3. With a common applied load and a constant crack depth the stress intensity factor for each of the six specimens is varied by varying the width of each specimen. After the NTO/MMH testing has been completed,

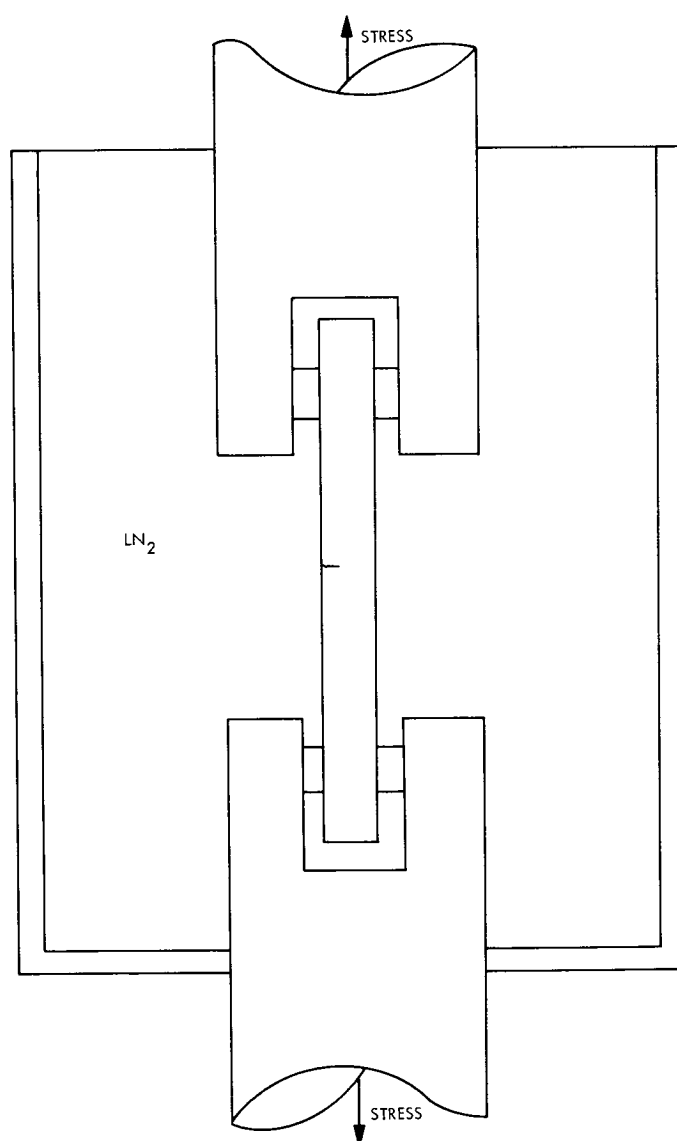


Fig. 2. Fracture toughness in proof environment

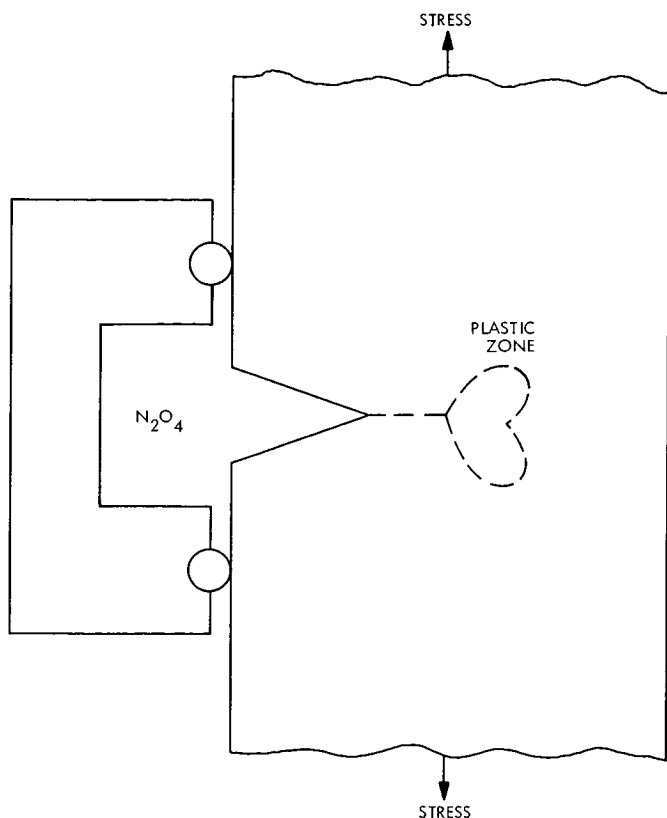


Fig. 3. Testing for crack growth threshold (1000 h)

the sustained load machines will be converted to  $\text{OF}_2$  and  $\text{B}_2\text{H}_6$ .

## 5. Test Plan

To optimize weight and reliability of a pressure vessel for long-term storage of stress corrosive fluids one needs to determine the trade-offs between  $K_{th}$ ,  $K_{Ic(pr)}$  and  $\sigma_{pr}$ . Since one seldom desires to proof test above the yield strength ( $\sigma_{ys}$ ) of the material,  $\sigma_{ys}$  can be used as a measure of  $\sigma_{pr}$ . This program will determine the variations in  $K_{th}$  and  $K_{Ic}$  as a function of the material yield stress for both fluid-material-environment combinations. Curves of  $K_{th}$  versus  $\sigma_{ys}$  and  $K_{Ic(pr)}$  versus  $\sigma_{ys}$  will be generated with 24-h tests. Longer duration tests up to 1000 h will be run on material of the optimum yield strength.

## C. Effect of Environment on Spacecraft Thermal Control Materials, W. Carroll and J. Crosby

### 1. Introduction

Degradation of thermal control coatings under combined conditions of ultraviolet radiation and vacuum is

initiated by photoproduct holes and electrons, either of which can undergo chemical reaction. Such chemical reactions change the structure of the coating, leading eventually to coloration. In this study,<sup>3</sup> the approach used to prevent optical degradation is to eliminate, in a coating system, those species produced by radiation that absorb significant quantities of thermal energy from the sun. This is accomplished by finding and introducing surface additives that act as recombinations centers, alternately capturing the holes and electrons, thus removing the photoproduct carriers with no net chemical change.

## 2. Results and Discussion

The fundamental portion of the investigation was the development of background knowledge for prediction of those chemical species which in the reduced form should capture holes and in the oxidized form should capture electrons. These studies have been brought to a satisfactory completion by use of the aqueous electrochemical technique. With the ZnO as the electrode, the cathodic current is directly proportional to the electron capture cross section of surface species. Various species were tested in the form of ions in the electrolyte. It was found that there was a maximum in the electron capture cross section for species with a surface state energy just below the bottom of the conduction band. The capture of holes by reducing agents (again introduced as ions in solution) was tested by a phenomenon termed "current doubling," which was observed in the following manner. With an indifferent electrolyte, the anodic current is solely due to photoproduct holes which reach the surface. However, with many two-equivalent species, the anodic current almost doubles because the unstable intermediate injects electrons into the conduction band. That is, captured holes will oxidize the ion in a one-equivalent process, but with a two-equivalent ion a highly unstable intermediate is formed. It decomposes spontaneously, injecting an electron to provide the second equivalent of oxidation. Thus for each hole reaching the surface, an electron is injected, and the current doubles. The phenomenon of current doubling can be used not only to determine when a species is oxidized in a two-equivalent process, but also to compare the hole capture cross section of one-equivalent and two-equivalent reducing agents.

In the studies of ZnO, several additive species potentially suitable as recombination centers were identified (all one-equivalent species with redox potentials, and hence surface energy levels, in a certain range). However, of the

<sup>3</sup>Performed by Stanford Research Institute under JPL Contract 951522.

species actually tested, the redox couple ferro/ferricyanide stood out as having optimum properties with respect to stability, ease of handling and high capture cross sections.

Tests of the effectiveness of the additive couple ferro/ferricyanide have been made on ZnO using two accelerated test procedures.

In one test procedure, vacuum photolysis of ZnO is monitored by measurement of the increase in dark conductance of ZnO crystals. This measurement permitted quantitative analysis of the action of the additive, and it was found that the rate of photolysis was decreased by as much as five orders of magnitude with a monolayer of a 1:1 mixture of ferro/ferricyanide additive on the surface. The analysis at this stage indicates that the iron cyanide behaves as a simple recombination center, as desired. From both experimental observations and theory it is clear that a mixture of both oxidation states must be present, and that the degradation rate varies inversely as the square of the total additive concentration.

According to the second test procedure, vacuum photolysis is monitored by electron spin resonance (ESR) of a signal at  $g = 1.96$  associated indirectly with donors in ZnO. It was shown for the test conditions that if the ferro/ferricyanide concentration exceeded about  $10^{-2}$  monolayers, no change in the ESR spectra occurred upon ultraviolet illumination in vacuum. The ESR study also provided some detailed information about the probable nature of the photodamage centers. For example, ultraviolet irradiation produces two resonance centers which exhibit behavior to oxygen, suggesting they may be associated with the optical damage centers giving rise to visible and infrared absorption. Also, from a temperature study it is concluded tentatively that some of the paramagnetic centers produced by photodamage are in the conduction band, and some may be associated with excess zinc arising from ZnO photolysis. Finally, the ESR measurements show that when dislocations are introduced in ZnO the characteristic resonance show greater susceptibility to photoirradiation.

Materials other than ZnO were examined, in particular  $ZrO_2$  and  $La_2O_3$ , with unsatisfactory results. With the electrochemical technique the fundamental studies indicated that it was necessary to determine only one parameter for each pigment, in order to suggest a redox couple suitable as a recombination center. This one parameter is the effective redox potential (energy level) of the conduction band edge of the semiconductor pigment. An attempt was made

to measure this parameter by a study of electron injection from strong one-equivalent reducing agents into the conduction band. By noting the redox couple (potential) where electron injection was no longer possible, the parameter of interest was determined. Such a measurement was entirely successful in ZnO studies. Unfortunately, the results obtained with a series of species on  $ZrO_2$  and  $La_2O_3$  could not be interpreted; injection apparently depends on parameters that could not be controlled. Thus it was concluded that another technique must be used to determine the effective redox potential of the conduction band for these two pigments.

### 3. Conclusions and Future Plans

During the previous report period, the work on photolysis of pigments and methods to prevent degradation was completed and the final report (Ref. 1) on this activity was submitted. While the program concept and approach has shown high potential productivity and technical merit, it is not planned currently to continue the work beyond this point.

Salient points of interest which have resulted from this effort and are of long-range significance to future studies of this nature are:

- (1) While it is apparent that several additive species potentially suitable as recombination centers in ZnO have been identified, it was not ascertained whether the additive is truly catalytic or if some depletion reaction is involved.
- (2) Attempts to apply the additive concept to several other pigment materials; namely,  $ZrO_2$  and  $La_2O_3$ , were unsuccessful. The reasons are not understood at this time, but electrical measurements aimed at selecting suitable recombination couples for these materials have produced anomalous results.
- (3) For ZnO, two distinct ESR signals associated with ultraviolet irradiation were identified, one of which is oxygen bleachable. At this point, correlation of this observation with known visible and infrared damage states has not been established.

### References

1. Morrison, S. R., Sancier, K. M., "Effect of Environment on Thermal Control Coatings," Final Report, SRI Project No. PAD-6146, Stanford Research Institute, Oct. 15, 1969, under JPL Contract No. 951522.

# XI. Applied Mechanics

## ENGINEERING MECHANICS DIVISION

### A. Optimization of Space Antenna Concepts,

J. N. Yang

#### 1. Simplification of Formulations

The antenna optimization concept (SPS 37-60, Vol. III, pp. 157-159) can be summarized as follows: Minimize the relative cost  $C^* = C/a_3$

$$C^* = \gamma D^{a_2} N^{a_5} b^{a_{10}} + N^{a_4} D^{a_4 a_6} \quad (1)$$

Subject to the constraints

$$G_1 = \eta \left( \frac{\pi D}{\lambda} \right)^2 \exp \left\{ - \left[ \frac{4\pi \delta(D, N, b)}{\lambda} \right]^2 \right\} \geq G_{01} \quad (2)$$

$$G_2 = \eta \left( \frac{\pi D}{\lambda + \Delta \lambda} \right)^2 \exp \left\{ - \left[ \frac{4\pi \delta(D, N, b)}{\lambda + \Delta \lambda} \right]^2 \right\} \geq G_{02} \quad (3)$$

$$\lambda \geq \lambda_0 \quad (4)$$

$$\Delta \lambda \geq \Delta \lambda_0 \quad (5)$$

$$W(D, N) \leq W_0 \quad (6)$$

in which

$$W(D, N) = ND^{a_6} \quad (7)$$

$$\delta^2(D, N, b) = \left( \frac{a_7 D^{a_8}}{b} \right)^2 + \left( \frac{a_{11} D}{N^{a_5}} \right)^2 + (a_{12} D)^2 \quad (8)$$

Where  $D, N, b, \lambda$ , and  $\lambda + \Delta \lambda$  are design variables that indicate antenna diameter, measure of weight per unit area, measure of manufacture precision, first operational wave length, and the second operational wave length, respectively.

In Eq. (1),

$$\gamma = \frac{a_1}{a_3} \quad (9)$$

is a measure of relative importance of the antenna cost itself and the cost of antenna payload. This parameter varies from mission to mission.

Equations (1-8) indicate that there are four design variables that are associated with the optimum design of an antenna operating at single frequency. One additional design variable is created that corresponds to one additional frequency operation. However, the available information indicates that  $a_s = 1$  and it can be shown that at optimum

$$\lambda = \lambda_0 \quad (10)$$

$$\Delta\lambda = \Delta\lambda_0 \quad (11)$$

Optimum design is realized when the first operational frequency is as high as allowable, and the second operational frequency is as close as allowable to the first operational frequency. The number of design variables is therefore reduced to three ( $D$ ,  $N$ , and  $b$ ), irrespective of the number of simultaneously operating frequencies.

## 2. Maximum Allowable Gain Constraint

In addition to the manufacturing tolerance  $\delta_m$  and the surface tolerance  $\delta_s$ , the antenna gain is limited by the inherent residue deformation  $\delta_c$  (temperature deformation, etc.). This deformation puts an upper bound on the antenna gain that cannot be attained even if  $\delta_m$  and  $\delta_s$  can be made infinitely small. This upper bound can be obtained from Eq. (3) by setting  $\delta_m = 0$ ,  $\delta_s = 0$ . For a single frequency operation condition,

$$G_1 = \eta \left( \frac{\pi D}{\lambda_0} \right)^2 \exp \left[ - \left( \frac{4\pi D a_{12}}{\lambda_0} \right)^2 \right] \quad (12)$$

with the derivative of Eq. (12) equal to zero with respect to  $D$ , Eq. (13) is obtained.

$$D = \frac{\lambda_0}{4\pi a_{12}} \quad (13)$$

Hence, the maximum gain attainable  $G_{1\max}$  is

$$G_{1\max} = \frac{\eta}{16a_{12}^2 e} \quad (14)$$

and the gain constant  $G_{01}$  should be less than  $G_{1\max}$ .

## 3. Numerical Optimization Technique

The objective function and the constraints given in Eqs. (1-8) are nonlinear functions of the design variables  $D$ ,  $N$ , and  $b$ . A gradient move technique is used in this discussion and is briefly described in the following: Consider a nonlinear objective function  $C$  of  $n$  design variables ( $A_1, A_2, \dots, A_n$ ) subject to  $m$  nonlinear constraints, i.e., minimize

$$C = C(A_1, A_2, \dots, A_n) \quad (15)$$

subject to

$$E_j = E_j(A_1, A_2, \dots, A_n) \leq E_{j0} \quad (16)$$

$$j = 1, 2, \dots, m$$

It is assumed that  $C$  is a monotonically increasing function of the design variables  $A_1, A_2, \dots, A_n$  so that at optimum, at least one of the constraints is active (i.e., the equality sign holds at least for one of the constraints in Eq. 16). This assumption is satisfied in the optimum antenna design as can be observed from Eq. (1).

Let  $\mathbf{V}$  and  $\mathbf{U}_j$  be the gradients of the objective function  $C$  and the constraint  $E_j$ , respectively,

$$\mathbf{V} = \nabla C = \sum_{k=1}^n \frac{\partial C}{\partial A_k} \mathbf{i}_k \quad (17)$$

$$\mathbf{U}_j = \nabla E_j = \sum_{k=1}^n \frac{\partial E_j}{\partial A_k} \mathbf{i}_k \quad j = 1, 2, \dots, m \quad (18)$$

in which both  $\mathbf{V}$  and  $\mathbf{U}_j$  are normal to  $C$  and  $E_j$  and  $\mathbf{i}_k$  is the unit vector in the positive direction of axis  $A_k$ .

**a. Phase I: steepest descent modification.** A design point  $B_1$  is first chosen arbitrarily in the acceptable domain defined by  $E_j \leq E_{j0}$  ( $j = 1, 2, \dots, m$ ) of the  $n$ -dimensional space  $A_1, A_2, \dots, A_n$ . The design is then modified by moving normal to the objective function  $C$  at  $B_1$ , i.e., in the direction of  $-\mathbf{V}$  at  $B_1$ , by a specified step from  $B_1$  to  $B_2$

with a reduction in  $C$ . This process is repeated until a constraint  $E_j = E_{j_0}$  is reached at point  $B_0$ . The direction of design modification  $-\mathbf{V}$  changes from point to point and has to be computed for each step of modification.

**b. Phase II: usable feasible direction modification.**  
Let  $\mathbf{Q}$  be a vector such that, at  $B_0$ ,

$$\mathbf{U}_j \cdot \mathbf{Q} \leq 0 \quad (19)$$

$$\mathbf{V} \cdot \mathbf{Q} \leq 0 \quad (20)$$

The direction  $\mathbf{Q}$  defines the so-called usable feasible direction. A systematic scheme for finding  $\mathbf{Q}$  (Ref. 1) is used in this discussion. The design point is modified from  $B_0$  along  $\mathbf{Q}$  in a specified step away from the constraint  $E_j = E_{j_0}$  into the acceptable domain with a reduction of objective function  $C$ , as shown by Ineqs. (19) and (20). The modification then proceeds along  $\mathbf{Q}$  at  $B_0$  until either one of two cases occur: (1) The objective function  $C$  starts to increase at the design  $B_2$ . Because the objective function  $C$  is nonlinear, the continuous modification of the design along  $\mathbf{Q}$  at  $B_0$  does not guarantee the monotonic decrease of  $C$  (it is true if  $C$  is linear). [Ineq. 19 does not imply that  $(\mathbf{V} \text{ at } B_2) \cdot (\mathbf{Q} \text{ at } B_0) \leq 0$ .] Should this situation occur, the steepest descent modification described in phase I is that employed at  $B_2$ . (2) A design point  $B_3$  on the constraint  $E_k = E_{k_0}$  is reached ( $k$  may be equal to  $j$ ), then another usable feasible direction  $\mathbf{Q}$  at  $B_3$  is computed, and the process of phase II is repeated until a design point  $B^*$  on the constraint is obtained at which the Kuhn-Tucker optimal condition is satisfied, i.e.,  $\mathbf{Q}$  cannot be found at  $B^*$  such that at least one of the inequality signs in Ineqs. (19) and (20) holds.

In the phase II procedure, the design point may be modified to a point  $B$  which is the intersection of many constraints, say constraints 1, 2, and 3. The usable feasible direction  $\mathbf{Q}$  at  $B$  is defined as (at  $B$ )

$$\left. \begin{aligned} \mathbf{U}_1 \cdot \mathbf{Q} &\leq 0 \\ \mathbf{U}_2 \cdot \mathbf{Q} &\leq 0 \\ \mathbf{U}_3 \cdot \mathbf{Q} &\leq 0 \\ \mathbf{V} \cdot \mathbf{Q} &\leq 0 \end{aligned} \right\} \quad (21)$$

and the Kuhn-Tucker optimality condition is satisfied at  $B$  if  $\mathbf{Q}$  cannot be obtained such that at least one of the inequality signs in Ineq. (21) holds.

The optimum design obtained is a local minimum and the global minimum can usually be obtained by choosing the minimum of local minima obtained from several different starting design points.

#### 4. Numerical Example

The optimum design of a rib-mesh-type parabolic antenna operating at either a single frequency or double frequencies is considered. The ground antenna requires that for a double frequency operation, the first operational frequency should be less than 10 GHz while the second operational frequency should be at least 3 GHz less than the first frequency. For a single frequency operation, the frequency should be less than 10 GHz. From past experiences and information (Refs. 2-4) it is reasonable to assume that  $a_2 = 2.0$ ,  $a_{10} = 1.0$ ,  $a_9 = 0.0$ ,  $a_6 = 2.0$ ,  $a_8 = 1.0$ , and  $\eta = 0.5$ . It is also assumed that  $a_4 = 1.0$ . The following parameters have been determined from an antenna design currently considered at JPL:  $a_7 = 0.14 \times 10^{-3}$ ,  $a_{12} = 0.08 \times 10^{-3}$ ,  $a_5 = 2.6$ , and  $a_{11} = 8.11 \times 10^{-7}$ . For instance, a 12.5-ft diam antenna will allow for a manufacturing tolerance  $\delta_m$  of 0.021 in. for a precision level  $b = 1.0$ , and a temperature and deployment tolerance  $\delta_c$  of 0.012 in. A 14-ft diam antenna with 30 ribs weighted 20.23 lb has a surface tolerance 0.05 in. while the same antenna with 20 ribs weighted 14.23 lb has a surface tolerance 0.125 in.

With this set of parameter values, the optimum antenna designs associated with the different gain constraints  $G_{01}$  and  $G_{02}$  (in this particular example,  $G_{01} = G_{02}$ ) are obtained and plotted in Figs. 1-4 for a specific value of  $\gamma$ . After the optimum values of  $D$ ,  $N$ , and  $b$  are obtained, the optimum manufacturing tolerance  $\delta_m$ , the optimum surface tolerance  $\delta_s$ , the optimum residue tolerance  $\delta_c$ , and the total tolerance  $\delta$  can be computed.

#### References

1. Zoutendijk, G., *Methods of Feasible Directions*, Elsevier Publishing Co., Amsterdam, 1960.
2. *Structures Technology for Large Radio and Radar Telescope System*, Edited by J. W. Mar and H. Liebowitz. Maple Press Co., 1969.
3. *Advanced Passive Communication Satellite Systems Comparison Studies: Vols. I-III*, Goodyear Aerospace Corp., Akron, Ohio, 1968.
4. *Space Erectable Large Aperture Reflector*, LMSC-A946613. Lockheed Missiles and Space Co., Sunnyvale, Calif., 1969.

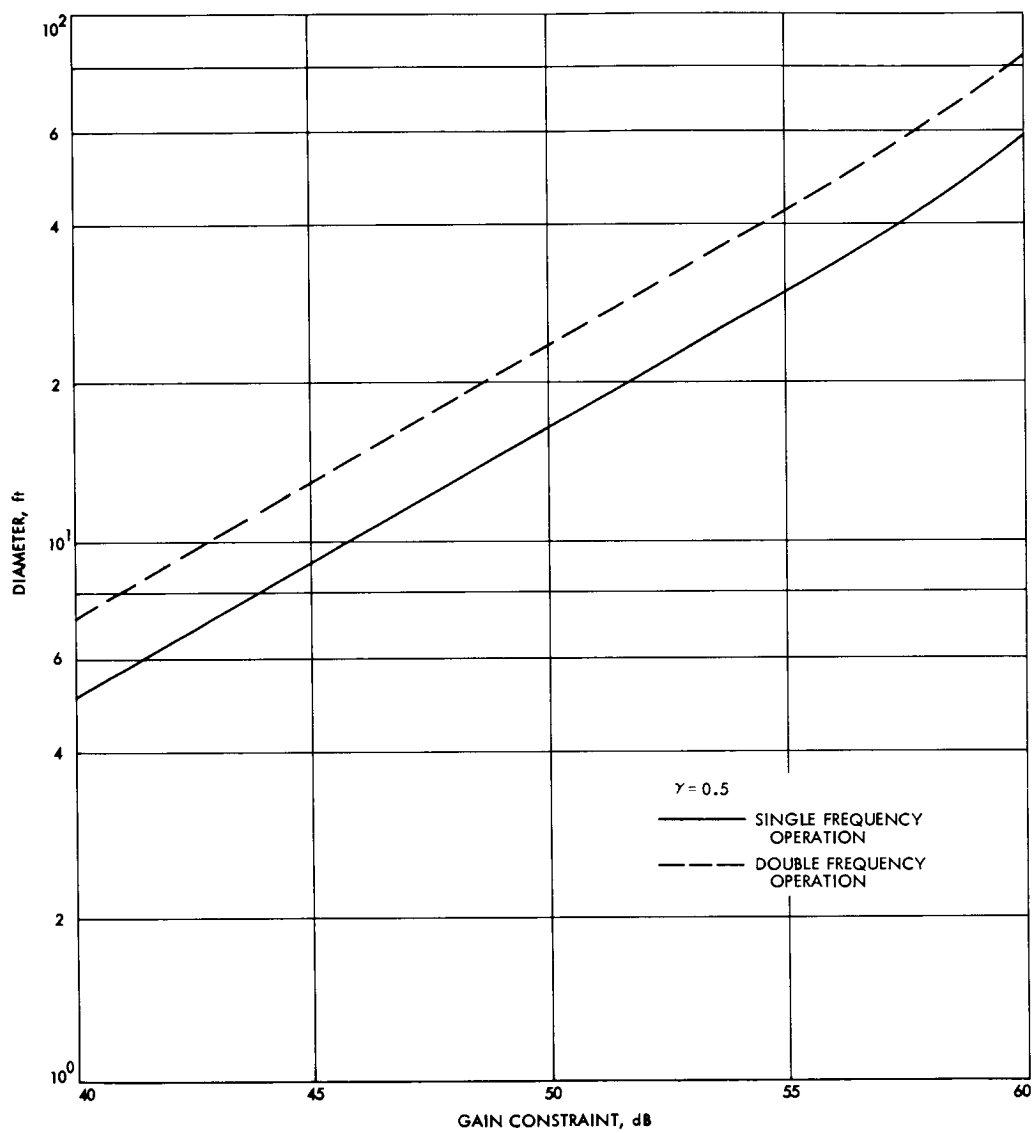


Fig. 1. Optimum antenna diameter vs gain constraints

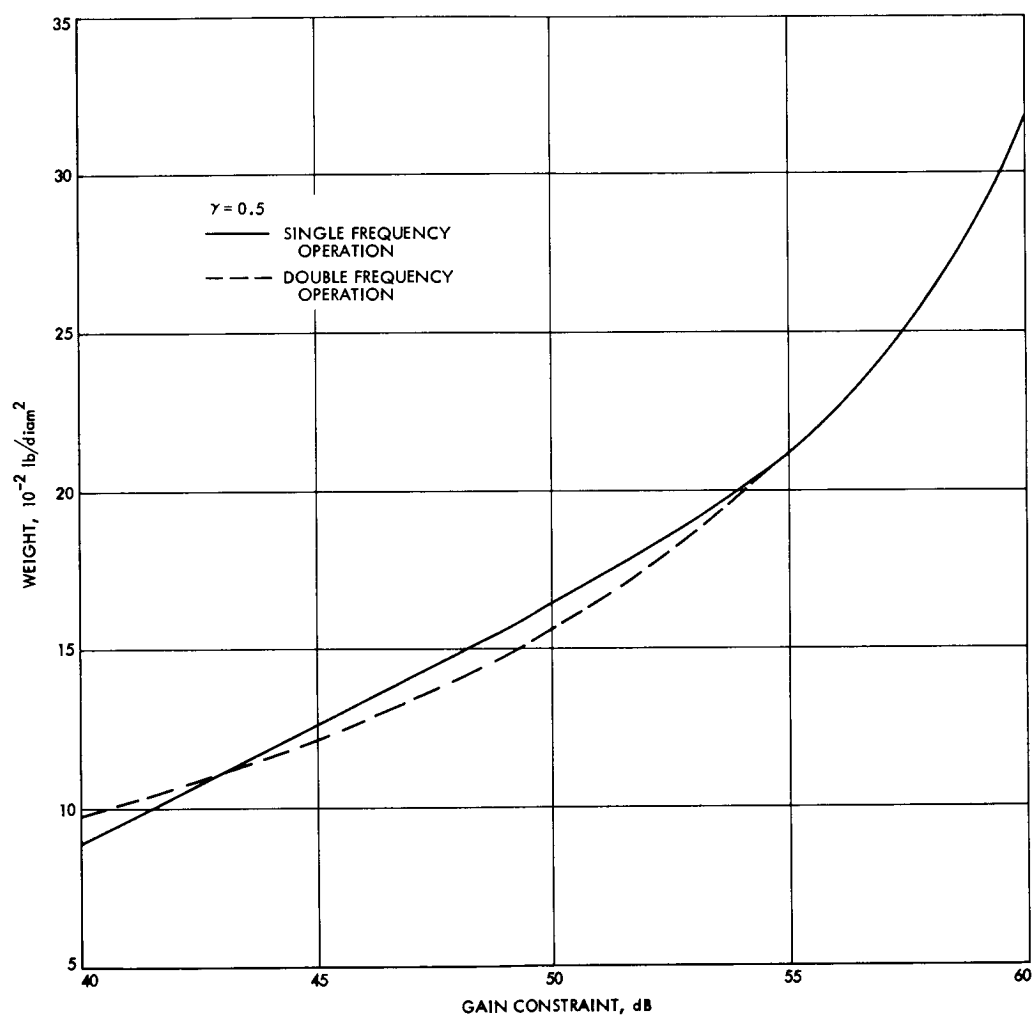


Fig. 2. Optimum weight per unit diameter vs gain constraints

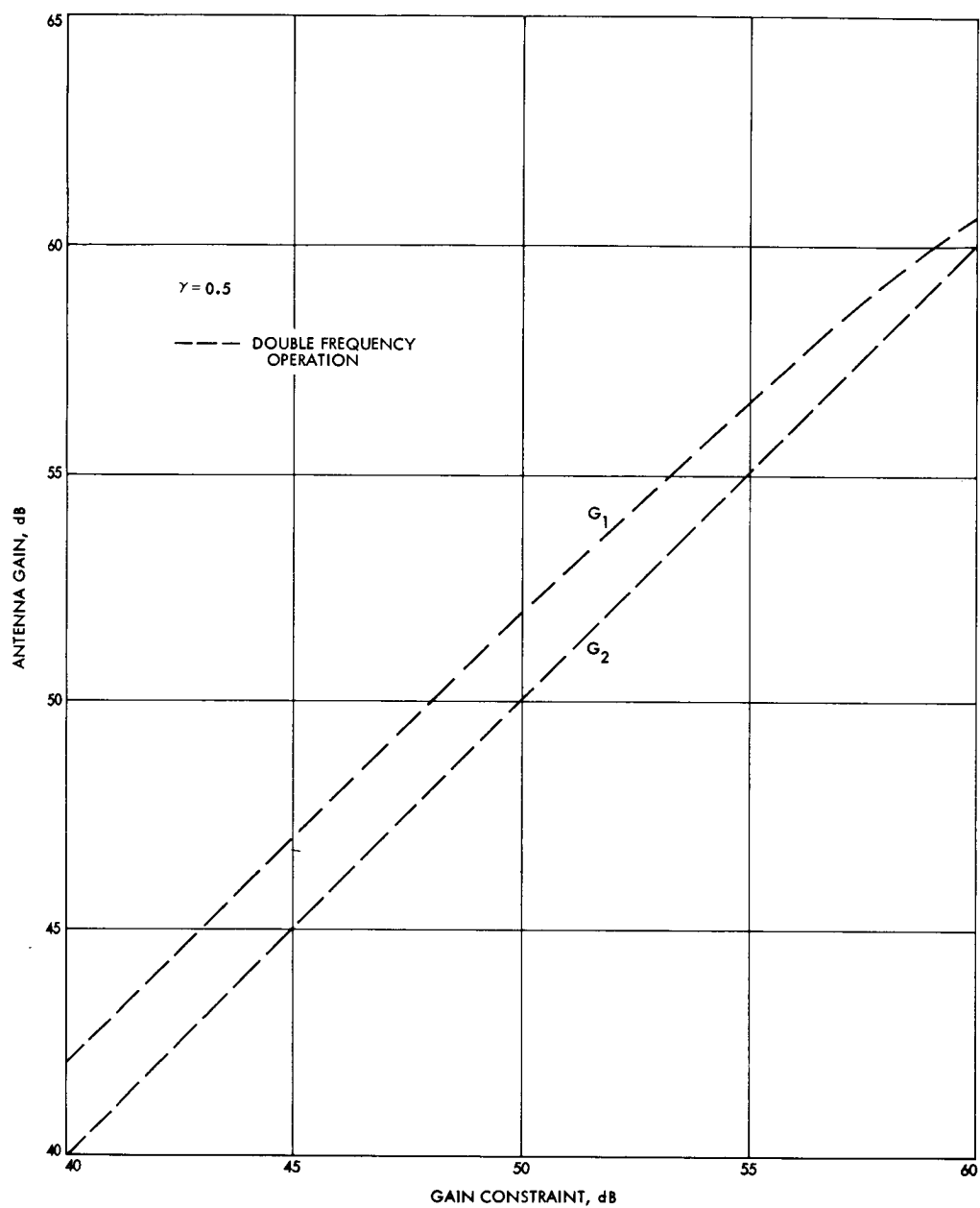


Fig. 3. Actual antenna gain vs gain constraints

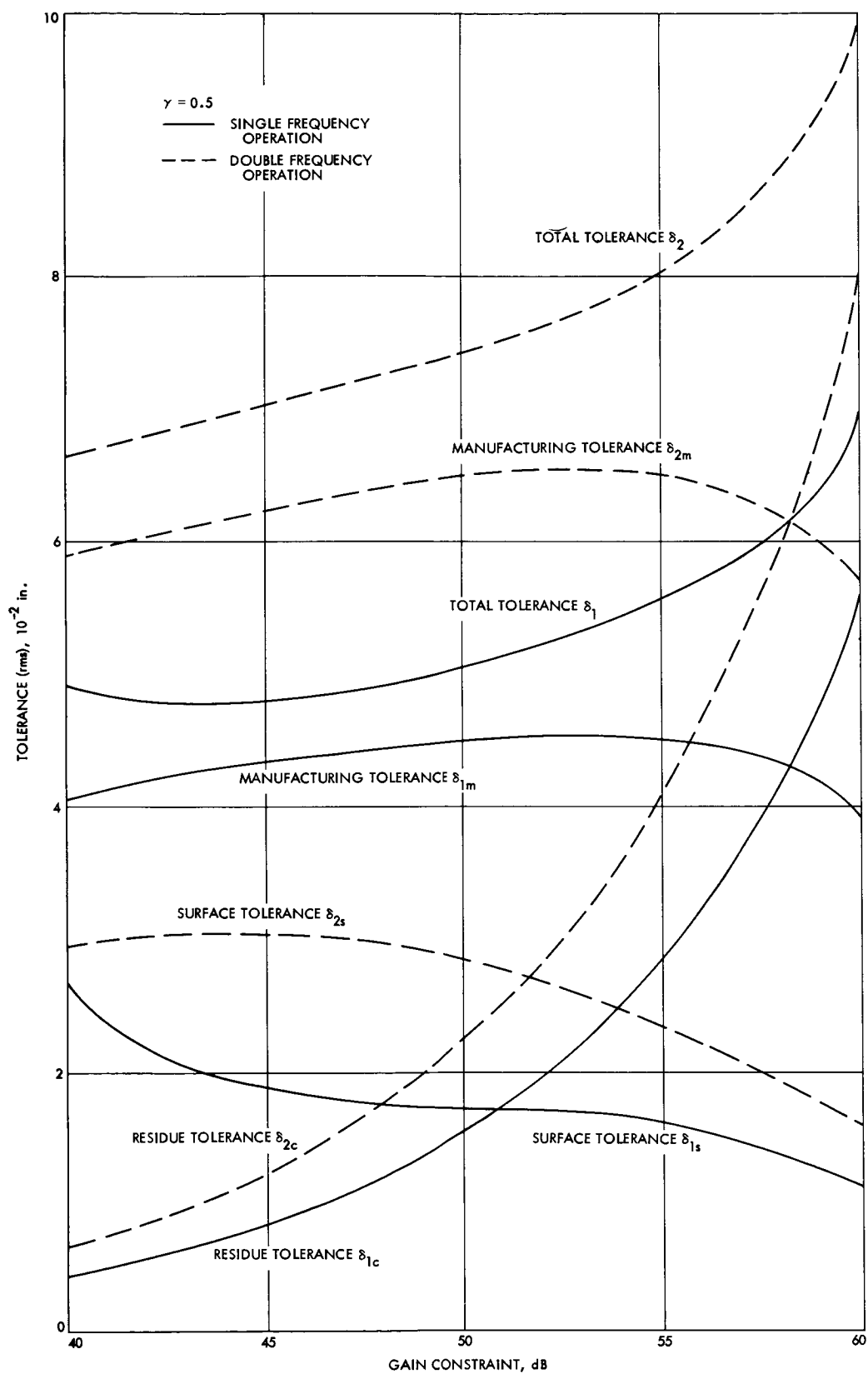


Fig. 4. Optimum rms tolerance vs gain constraints

## B. Receptance Coupling Program, R. D. Simpson

### 1. Introduction

The method of determining the dynamic response of a system from the subsystems characteristics has been previously explored. The basic problem considered was that of joining together and determining the dynamic response of a number of subsystems that may simulate, for example, a launch vehicle, a spacecraft, an entry capsule, and a landing system. It was shown how the system receptance matrix (transfer matrix), the frequency response vector, and the crosspower spectral-density matrix of the coupled system can be obtained from analytically and/or experimentally determined receptances of the component systems.

### 2. Discussion

The Receptance Coupling Program (RECEP), which implemented the above method, was discussed in

SPS 37-49, Vol. III, pp. 159-161 and SPS 37-56, Vol. III, pp. 164-165. Recent activities have been twofold:

- (1) Conversion to the Univac 1108 system and utilization of some of the unique features available with system.
- (2) Additional developments in RECEP, such as improved internal subsystem interfacing and program flexibility.

Because conversion to the Univac 1108 system meant recoding all assembly language subroutines to Fortran V, the opportunity was taken to restructure RECEP. The executive portion of the program has been completely restructured. As shown in Fig. 5, the executive has been redesigned for nonstationary as well as deterministic and stationary input.

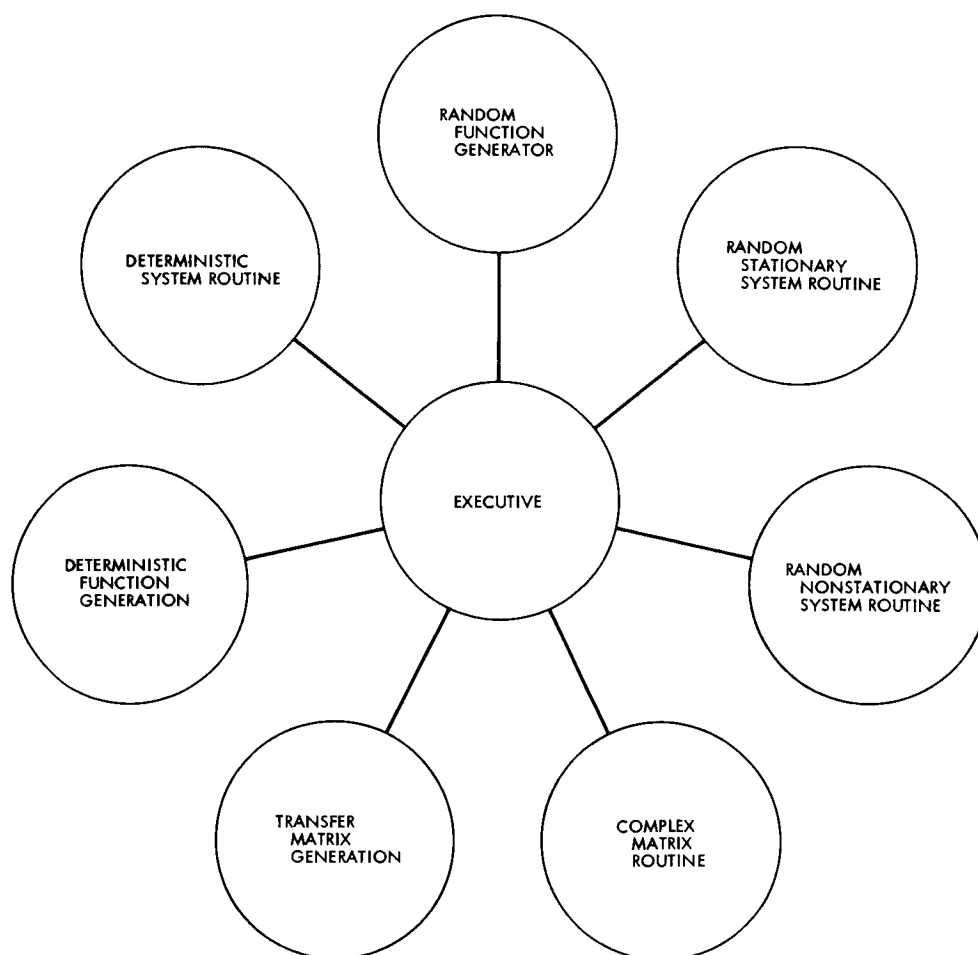


Fig. 5. Block diagram of redesigned RECEP executive program

The transfer matrix generation contains analytical and experimental routines for computing the complex matrix **H**. Currently this area accepts only SAMIS-type data; however, analytical generation need not depend on SAMIS-type output. It could, for example, use a NASTRAN-type output. Currently, the RECEP routine for reading a NASTRAN tape has not been written, but is prepared to accept one. However, the experimental IBM 7094 transfer matrix routine is available and can be converted to the Univac 1108 if needed. Additionally, the user can request a mixture of SAMIS, NASTRAN, and experimental data in each run.

The new complex matrix routine will allow for arithmetical transfer matrix operation. This routine can be useful when the coupling routine is not desired in some matrix manipulation, as in the study of closed-loop frequency response characteristics.

The deterministic function generator accepts analytically derived functions  $f(t)$ ,  $F(w)$ , or converted analog telemetry data  $f(t)$ . Currently, the analytical function routine accepts tape or card input for either the time history  $f(t)$  or its Fourier transform  $F(w)$ . The analytical tape is assumed to be in a format compatible with RECEP. The analog telemetry data is digitized by the Data Analysis Facility (DAF) and transferred to a tape. This tape is usually referred to as a DAF tape. The DAF tape routine is more flexible than previously reported. It allows for file skip, data-record skip, data skip, signal offset, and calibration correction factor options. The input requirements for this routine have been sharply reduced. The transfer matrix routine for the experimental receptance is in the same DAF format.

Occasionally an undesired trend appears in the experimental data and should be removed. Hence, a trend analysis subroutine is being added that uses the least-squares procedure. A low-degree (first through fourth degree) polynomial curve fit to the data will be allowed. This polynomial is evaluated for each sample and subtracted from the data. If the polynomial has been chosen appropriately, the trend will be removed with minimal effect on the information of interest. There is also an option available that simply gives the data an ordinate shift.

### 3. Concluding Remarks

The deterministic system routine can now handle single and multiple component problems as well as the inverse problem. The inverse problem considers that the re-

sponse of the system is known and that the excitation is required. This procedure is presently being used on the *Mariner Mars 1971* Project. Currently, with a possible exception in the complex matrix routine, matrix operations that require inversions are accomplished via the standard matrix decomposition procedure available through the JPL program library. The discussion of the random process capability of RECEP is a subject for future discussion. For nonstationary type problems, a double Fourier transform is discussed in SPS 37-59, Vol. III, pp. 168-171. This routine, which will do double as well as single transformations, will be added to RECEP in the future.

### Reference

Heer, E., "Coupled Systems Subjected to Determinate and Random Input," *Int. J. Solid Struct.*, Vol. 3, pp. 155-166, 1967.

## C. On the Maximum Dynamic Response of Structures, M. Shinozuka

### 1. Introduction

This article discusses an analytical approach previously applied to earthquake studies<sup>1</sup> and an extension of this approach that possesses potential usefulness in the dynamic environmental test with a shaker as well as dynamic analysis of spacecraft. Although the discussion in this article emphasizes analytical aspects of the problem, its practical implications are self-explanatory.

The one-dimensional (one-component) problem is emphasized at this time. However, a straightforward extension of the approach to a multi-dimensional input situation is also presented. The existence of all the integrals and derivatives appearing in the following discussion is assumed.

### 2. Analysis I

Consider a system represented by a linear time-invariant operator  $L$  that relates input  $x(t)$  and output  $y(t)$  in the form

$$L[y(t)] = x(t) \quad (1)$$

<sup>1</sup>Shinozuka, M., *On the Maximum Structural Response to Earthquake Acceleration*, Technical Report 5 (NSF GK-3858). Department of Civil Engineering and Engineering Mechanics, Columbia University, New York (in press *J. Eng. Mech.*, Proc. ASCE).

and possesses the impulse response function  $h(t)$  and the frequency response function  $H(\omega)$ . Writing  $X(\omega)$  and  $Y(\omega)$  for the Fourier transforms of  $x(t)$  and  $y(t)$ , respectively, the response  $y(t)$  can be written as

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) X(\omega) e^{i\omega t} d\omega \quad (2)$$

where  $i$  is the imaginary unit.

In the present discussion,  $x(t)$  and  $y(t)$  can be displacement, acceleration force, etc., but they do not necessarily represent the same quantity.

In accordance with Churchill (Ref. 2)

$$\left| \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) H(\omega) e^{i\omega t} d\omega \right| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)| |H(\omega)| d\omega = I \quad (3)$$

and also with the aid of the Schwartz inequality

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)| |H(\omega)| d\omega \leq \left( \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega \right)^{1/2} \left( \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega \right)^{1/2} \quad (4)$$

The combination of Eqs. (2) and (3) and Eq. (4) yields

$$\max_t |y(t)| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)| |H(\omega)| d\omega \leq \left( \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega \right)^{1/2} \left( \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega \right)^{1/2} \quad (5)$$

where  $\max_t |y(t)|$  indicates that the maximum absolute value of  $y(t)$  for  $-\infty < t < \infty$ .

The relationship involving the first and the last member of Eq. (5) was used by Drenick (Ref. 1) leading to the following conclusion. By virtue of Parseval's theorem,

$$\int_{-\infty}^{\infty} x^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega \quad (6)$$

is obtained, which states that the integral of  $|X(\omega)|^2/2\pi$  over the entire  $\omega$  domain is equal to the total energy associated with the input  $x(t)$ . Consider a class of input functions with total energy  $\leq M^2$ :

$$\int_{-\infty}^{\infty} x^2(t) dt \leq M^2 \quad (7)$$

Then, due to Eq. (5), an upper bound of a set of maximum absolute responses produced by such a class of inputs is  $MN$ ;

$$\max_t |y(t)| \leq MN \quad (8)$$

where

$$N = \left( \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega \right)^{1/2} \quad (9)$$

A least-favorable input under the condition of Eq. (7) is then given by  $x_0(t)$  with the Fourier transform  $X_0(\omega)$  satisfying

$$X_0(\omega) = \frac{M}{N} H^*(\omega) \quad (10)$$

where the asterisk indicates the complex conjugate. The fact that  $x_0(t)$  is truly a least-favorable input can be seen by substitution of Eq. (10) into Eq. (2) in which  $t$  is set equal to zero

$$y(0) = \frac{M}{2\pi N} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega \quad (11)$$

It is realized that this is equal to the maximum possible value  $MN$  the absolute response can reach. Application of the Fourier inverse transform to  $X_0(\omega)$  in Eq. (10);

$$\begin{aligned} x_0(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X_0(\omega) e^{i\omega t} d\omega = \frac{M}{2\pi N} \int_{-\infty}^{\infty} H^*(\omega) e^{i\omega t} d\omega \\ &= \frac{M}{N} h(-t) \end{aligned} \quad (12)$$

Therefore, the conclusion, in accordance with Ref. 1, is that the least-favorable input is the mirror image of the impulse response function  $h(t)$  of the system with

respect to  $t = 0$  multiplied by a factor  $M/N$ . The corresponding least-favorable absolute response is  $MN$  occurring at  $t = 0$ .

### 3. Analysis II

A clear picture of the maximum possible value of the response of a known system when the upper bound of the total energy of the input can be specified is given in Ref. 1. It is not unusual that a sufficient amount of information is available on which the Fourier transform  $X(\omega)$  (of the input) and its related quantities can be constructed from observations in a manner considered reasonable from an engineering standpoint. The envelope of  $|X(\omega)|$  appears to be particularly important in this connection, since a system is usually subjected to a number of inputs that are similar but not identical. A reasonable value for a safety factor may be introduced in the process of constructing an envelope. Let  $|X_e(\omega)|$  represent such an envelope. Then, from Eq. (5), the response has an upper bound

$$I = \int_{-\infty}^{\infty} |X_e(\omega)| |H(\omega)| \frac{d\omega}{2\pi}$$

which is, in turn,  $\leq MN$  where  $M$  and  $N$  are again defined by Eqs. (7) and (9), respectively, with  $|X(\omega)|$  replaced by  $|X_e(\omega)|$ . This is compatible with the expectation that an increased amount of information, in this case, knowledge of  $|X_e(\omega)|$  rather than

$$\int_{-\infty}^{\infty} |X_e(\omega)|^2 d\omega$$

should produce a better upper bound.

For example, consider a case where  $|X_e(\omega)| = X_0 = \text{constant}$  ( $-\Omega \leq \omega \leq \Omega$ , where  $\Omega$  is a positive constant). This may be assumed for the envelope of the modulus of the Fourier transform of transient torsional acceleration caused by a booster engine cutoff and transmitted through the launch vehicle and adapter to the spacecraft (see Figs. 13 and 15 of Ref. 3). Then

$$\begin{aligned} \max_t |y(t)| &\leq \frac{1}{2\pi} \int_{-\Omega}^{\Omega} |X_e(\omega)| |H(\omega)| d\omega \\ &= \frac{X_0}{2\pi} \int_{-\Omega}^{\Omega} |H(\omega)| d\omega \leq MN \end{aligned} \quad (13)$$

Consider further, a simple torsional vibratory system with an angular response  $y(t)$  relative to the base can be given by

$$\ddot{y}(t) + 2\zeta\omega_0\dot{y}(t) + \omega_0^2 y(t) = -x(t) \quad (14)$$

where  $x(t)$  is the angular acceleration of the base. Then, the frequency response function of  $\dot{y}(t)$  is

$$H(\omega) = \frac{-i\omega}{(\omega_0^2 - \omega^2) + 2i\zeta\omega\omega_0} \quad (15)$$

where the relative angular velocity  $\dot{y}(t)$  is considered as the response of major interest. Then, the last integral in Eq. (13) can be shown to become

$$\int_0^{\Omega} |H(\omega)| d\omega \approx \ln \left[ \left( \frac{\Omega}{\omega_0} \right)^2 \frac{1}{\zeta^2} \right] \quad (16)$$

where the approximation is valid for  $\Omega/\omega_0 \gg 1$  and  $\zeta \ll 1$ .

If, for example,  $\Omega = 10\omega_0$ , Eq. (16) is 10.59 and therefore  $I = 1.59X_0$ :

$$\max_t |\dot{y}(t)| \leq 1.59X_0 \quad (17)$$

If, however, one uses Eq. (8) with  $M^2 = \Omega X_0^2/\pi$  and

$$N^2 = \frac{1}{\pi} \int_0^{10\omega_0} |H(\omega)|^2 d\omega = \frac{4.97}{\omega_0} \quad (18)$$

then

$$\max_t |\dot{y}(t)| \leq 3.98 X_0 \quad (19)$$

This result indicates that Eq. (17) provides a better upper bound ( $\approx 40\%$  of the value given by Eq. 19) for the present example.

Once

$$\max_t |\dot{y}(t)|$$

is established, maximum absolute relative response

$$\max_t |y(t)|$$

and maximum absolute acceleration

$$\max_t |\ddot{y}(t) + x(t)|$$

can be estimated in approximation by

$$\max_t |y(t)| \approx \frac{\max_{\omega_0} |y(t)|}{\omega_0} \quad (20)$$

$$\max_t |y(t) + x(t)| \approx \omega_0 \max_t |y(t)| \quad (21)$$

Although closed-form integrations or their approximate values were found for  $M$ ,  $N$ , and  $I$  in the example described in Eqs. (20) and (21), these integrations generally have to be evaluated numerically.

The relationship between the envelope and the moduli of Fourier transform of those inputs that are considered to represent the disturbance originated at the same source warrants consideration. Unless sufficient evidence requires a more rigorous treatment, it is not unreasonable to determine the envelope on the basis of an engineering judgment. For example, if two inputs are somehow measured and their moduli  $|X_1(\omega)|$  and  $|X_2(\omega)|$  are computed as shown in Fig. 6, an envelope  $|X_e(\omega)|$  is obtained as line A covering those moduli at any point of frequency. Line B represents an alternative envelope that might be interpreted as having a larger safety factor. The fact that  $|X_e(\omega)| \geq |X_i(\omega)|$  can evidently be used in Eq. (3) to show that the upper bound of the response associated with  $|X_e(\omega)|$  is larger than those with  $|X_i(\omega)|$ .

#### 4. Discussion of Least-Favorable Input Under the Given Modulus Envelope

The least-favorable input  $x_0(t)$  (for the general case) defined as the input that produces the maximum absolute response equal to

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |X_e(\omega)| |H(\omega)| d\omega$$

under the specified envelope  $|X_e(\omega)|$  of the modulus of Fourier transform of the input. Express  $X_0(\omega)$  (Fourier transform of  $x_0(t)$ ) and  $H(\omega)$  in their polar forms:

$$\left. \begin{aligned} H(\omega) &= |H(\omega)| e^{i\theta(\omega)} \\ X_0(\omega) &= |X_0(\omega)| e^{i\phi(\omega)} \end{aligned} \right\} \quad (22)$$

and choose  $|X_0(\omega)|$  and  $\phi(\omega)$  such that  $|X_0(\omega)| = |X_e(\omega)|$  and  $\phi(\omega) = -\theta(\omega)$ . Then

$$y(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_0(\omega) H(\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X_e(\omega)| |H(\omega)| d\omega \quad (23)$$

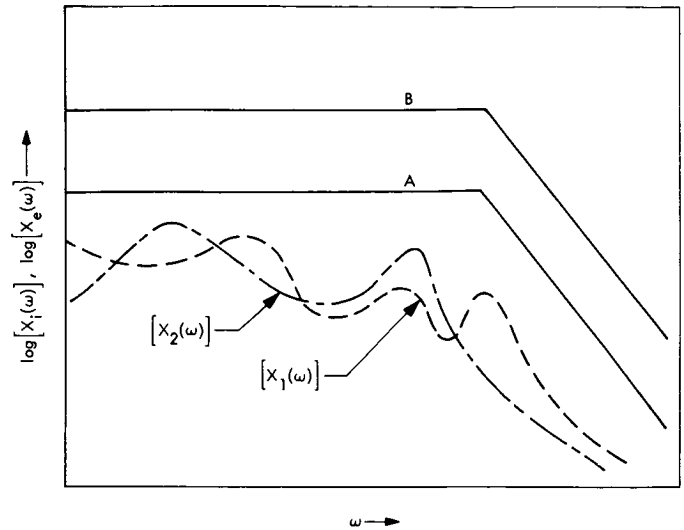


Fig. 6. Schematic diagrams of  $|x_i(\omega)|$  and  $|X_e(\omega)|$

Equation (23) indicates that the maximum absolute value that the response can possibly reach is attained at  $t = 0$  if the following input  $x_0(t)$  is applied to the system:

$$\begin{aligned} x_0(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |X_e(\omega)| e^{-i\theta(\omega)} e^{i\omega t} d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|X_e(\omega)|}{|H(\omega)|} H^*(\omega) e^{i\omega t} d\omega \end{aligned} \quad (24)$$

An implication of Eq. (24) for shaker testing is that the response at  $t = 0$  of a spacecraft to  $x_0(t)$  given in Eq. (24) is the least favorable response that ever has to be considered under the given envelope  $|X_e(\omega)|$ .

Unfortunately, Eq. (24) generally cannot be integrated in closed form. Therefore, numerical Fourier (one-dimensional) inversion to obtain  $x_0(t)$  is suggested, particularly since such a numerical procedure has become almost routine in the advent of fast Fourier transform.

If  $|X_e(\omega)| \cdot H^*(\omega) / |H(\omega)|$  cannot, but  $|X_e(\omega)| / |H(\omega)|$  can be inverted in closed form, the following result based on the convolution theorem may be useful.

$$x_0(t) = \int_{-\infty}^{\infty} h(-t + \tau) f(\tau) d\tau = \int_{-\infty}^{\infty} h(-\tau) f(t - \tau) d\tau \quad (25)$$

where  $f(t)$  is the inverse Fourier transform of  $|X_e(\omega)| / |H(\omega)|$ .

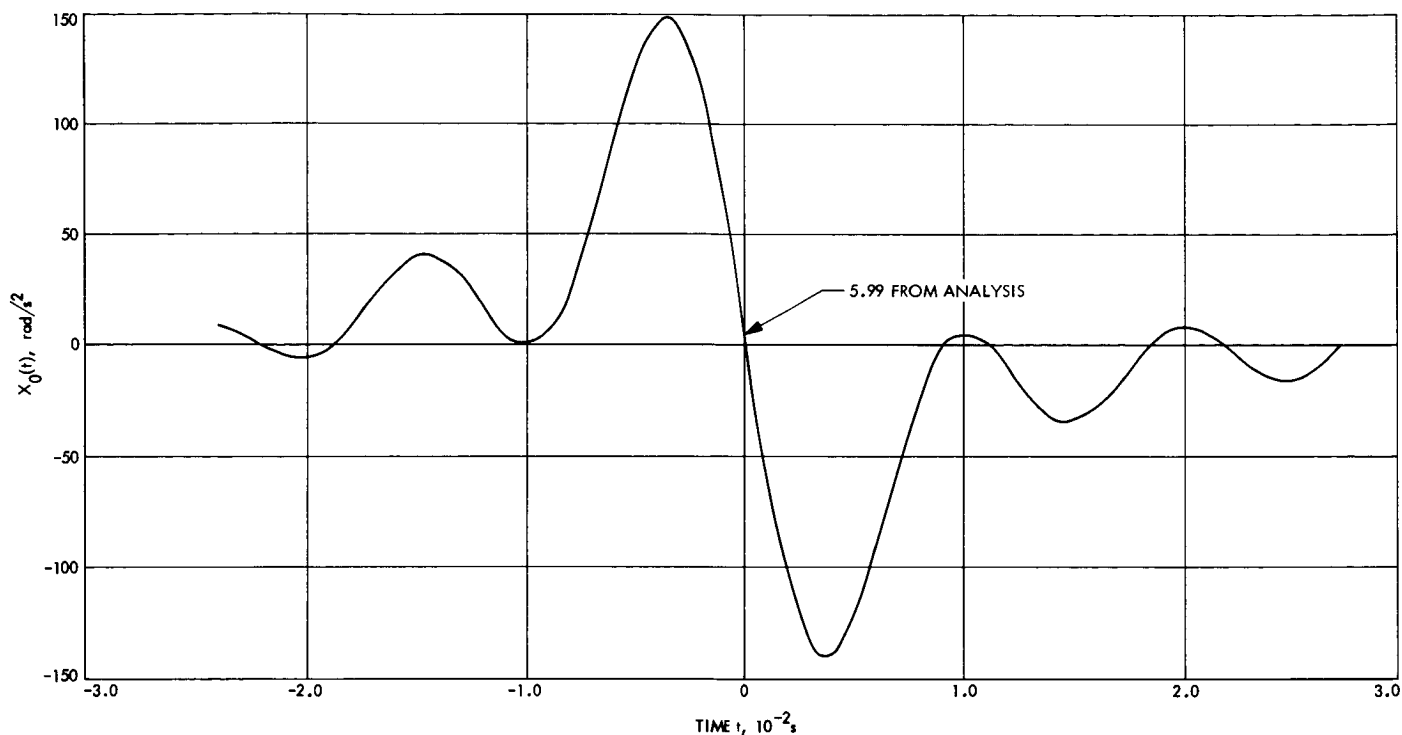


Fig. 7. A least-favorable input

As an example (Fig. 7), the least-favorable input  $x_0(t)$  is obtained by means of numerical integration of Eq. (24) for the case where  $H(\omega)$  is of the form given in Eq. (15) with  $\omega_0 = 10\pi$  rad/s and  $\zeta = 0.05$  and  $|X_e(\omega)| = X_0$  (constant) in  $-200\pi < \omega < 200\pi$ .

## 5. Multi-Dimensional Excitation

The preceding discussion can be applied to multi-dimensional excitation situations with a straightforward modification. Suppose that response  $y(t)$  and excitation  $x(t)$  are column vectors of dimensions  $m$  and  $n$ , respectively. Then, Eq. (2) is still valid if  $X(\omega)$  represents the Fourier transform of the column vector  $x(t)$  and  $H(\omega)$  the frequency response function matrix with  $H_{ij}(\omega)$  being the frequency responses of  $y_i(t)$  to  $x_j(t) = e^{i\omega t}$  ( $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ ). Therefore,

$$\max_t |y_i(t)| \leq \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{j=1}^n H_{ij}(\omega) X_j(\omega) e^{i\omega t} d\omega \right| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \sum_{j=1}^n H_{ij}(\omega) X_j(\omega) \right| d\omega \quad (26)$$

Equation (26) indicates that an upper bound for the maximum value of  $|y_i(t)|$  can be established by constructing an envelope for

$$\left| \sum_{j=1}^n H_{ij}(\omega) X_j(\omega) \right|$$

## References

1. Drenick, R. F., "Functional Analysis of Effects of Earthquakes," collection of preprinted papers presented at the Second Joint United States-Japan Seminar in Applied Stochastics, Vol. 1, Washington. Sept. 19-24, 1968.
2. Churchill, R. V., *Introduction to Complex Variables and Applications*, p. 79. McGraw-Hill Book Co., Inc., New York, 1948.
3. Heer, E., and Trubert, M. R., *Analysis of Space Vehicle Structures Using the Transfer Function Concept*, Technical Report 32-1367. Jet Propulsion Laboratory, Pasadena, Calif., Apr. 1, 1969.

## D. Deformation Modes of a Simple Rotating Structure, E. O. Weiner

### 1. Introduction

Deformation modes and natural frequencies are obtained for a simple rotating structure with lumped masses. The structure is constrained in a manner appropriate to the hybrid-coordinate method of analysis of Ref. 1. Having the relative motion equations that allow for centrifugal and Coriolis effects as well as the shift in center of mass, the linear equations of motion are completed by relating force and displacement differentials at a loaded state corresponding to the steady rotation of the structure. A brief description of the matrix algebra involved with the solution of the equations of motion is presented.

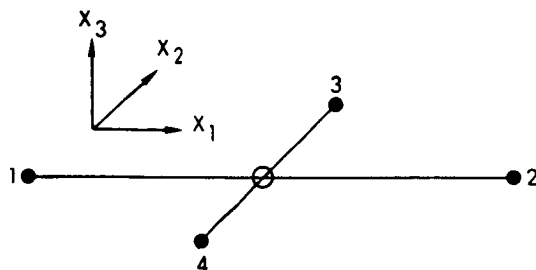
### 2. Definition of Problem<sup>2</sup>

The structure consists of two uniform and weightless beams bisecting one another at right angles. The intersection is attached to a 4-lb point mass, and each beam end carries a 1-lb point mass.

Each beam has its cross-section principal axes parallel to coordinate lines. The stiffnesses for each half beam considered cantilevered from the intersection are given tabulated below.

Beam	In-plane bending, lb/ft	Out-of-plane bending, lb/ft	Extension, lb/ft
Long	1.0	4.0	1.226
Short	9.0	25.0	11.034

The beam intersection is permitted to translate freely, but the beam intersection must remain parallel to an inertial plane and rotate with a constant spin. The  $x_1$  and  $x_2$  coordinate axes (shown below) are to be considered attached to the beam intersection.



<sup>2</sup>This example was suggested by P. W. Likins.

### 3. Equations of Motion

All displacements are measured in the  $x_1x_2x_3$  frame. Because of the symmetry of the structure, there is a steady motion that involves constant and equal but opposite radial displacements of diametrically opposed end masses. The term  $u_{ij}$  is the displacement of the  $i$ th mass along the  $x_j$  coordinate line measured from the position of the mass in the steady motion.

The equations of motion for the displacements  $u_{ij}$  are derived in Ref. 1. They may be written

$$M'\ddot{u} + G'\dot{u} + C'u = f \quad (1)$$

Here

$$u = (u_{11}, u_{12}, u_{13}, u_{21}, \dots, u_{43})^T$$

and

$$f = (f_{11}, f_{12}, f_{13}, f_{21}, \dots, f_{43})^T$$

represents the forces on the end masses in addition to those of the steady motion. The  $12 \times 12$  matrices  $M'$ ,  $G'$ , and  $C'$  are those displayed in Ref. 1, Figs. 5-7, respectively, with rows and columns corresponding to angular displacements deleted. The appropriate choices for symbols found in these figures are

$$m^1 = m^2 = m^3 = m^4 = \frac{1.0}{32.2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathcal{M} = \frac{8.0}{32.2}$$

$$\tilde{\Omega}^a = \tilde{\Omega}^a = 0$$

$$(\tilde{C}_\omega) = \begin{bmatrix} 0 & -\Omega & 0 \\ \Omega & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where  $\Omega$  is the constant spin of the beam intersection.

### 4. Stiffness Matrix

Using beam-column theory,  $f$  is nonlinearly related to  $u$ . In order to set up a linear eigenvalue problem from

the equations of motion, the derivative of the force-displacement relation evaluated at the loaded state of the steady motion is used to write

$$f = -Ku \quad (2)$$

where  $K$  is the stiffness matrix. It is diagonal, and its elements corresponding to extension are those of the tabulation given in *Subsection 2*. Each bending element, however, is the corresponding element  $k_0$  of that tabulation increased to

$$k_t = \frac{k_0}{\frac{3}{(\beta l)^2} \left( 1 - \frac{1}{\beta l} \tanh \beta l \right)} \quad (3)$$

where

$$(\beta l)^2 = 3 \frac{k_a}{k_0} \cdot \frac{m\Omega^2}{k_a - m\Omega^2} \quad (4)$$

Here  $k_a$  is the extensional stiffness and  $m$  is the end mass of the half-beam in question. The increase in bending stiffness can be attributed to the presence of the centrifugal forces of the steady motion.

## 5. Linear Eigenvalue Problem

By combining Eqs. (1) and (2), the equations of motion take the form

$$M'\ddot{u} + G'\dot{u} + (C' + K)u = 0 \quad (5)$$

where  $M'$  is positive definite and symmetric,  $G'$  is skew-symmetric, and  $C' + K$  is positive definite for sufficiently small  $\Omega$  and symmetric. The solutions to these equations were obtained by using a particular set of SAMIS control cards with the matrices  $M'$ ,  $G'$ , and  $C' + K$  as data.<sup>3</sup> Only a brief description of the matrix algebra involved is given here.

The eigenvalue problem

$$[M' - \lambda(C' + K)] \xi = 0$$

is solved by methods similar to those given in Ref. 2. Letting  $\Gamma \lambda \downarrow$  be the diagonal matrix of eigenvalues and letting  $\Phi$  be the corresponding matrix of eigenvectors normalized with respect to  $M'$ , the transformation

$$u = \Phi v$$

<sup>3</sup>The control cards were provided by J. Chisholm.

puts Eq. (5) in the form

$$\ddot{v} + \Phi^T G' \Phi \dot{v} + \Gamma \lambda \downarrow v = 0$$

This set of equations is equivalent to the following set of first order ordinary differential equations:

$$iA\dot{q} + Bq = 0 \quad (6)$$

Here,

$$A = \frac{1}{i} \begin{bmatrix} 0 & -I \\ I & \Phi^T G' \Phi \end{bmatrix}$$

$$B = \begin{bmatrix} I & 0 \\ 0 & \Gamma \lambda \downarrow \end{bmatrix}$$

$$q = \begin{bmatrix} \dot{v} \\ v \end{bmatrix}$$

where both  $A$  and  $B$  are nonsingular and hermitian while  $B$  is positive definite. By assuming a solution of the form

$$q = \eta e^{i\omega t}$$

Eq. (6) may be reduced to finding the eigenvectors and eigenvalues of a nonsingular hermitian matrix. By equating real and imaginary parts, this problem is equivalent to one of a symmetric matrix of double the order. Each solution, expressed in the original coordinates, may be put in the form

$$u = (\xi_{11}, \xi_{12}, \dots, \xi_{43})^T e^{i\omega t} \quad (7)$$

where elements  $\xi_{ij}$  are complex in general, and where  $\omega$  is the natural frequency of the mode. Both the real and imaginary parts of Eq. (7) are solutions to Eq. (5). The real part may be written

$$Re\, u = \begin{bmatrix} |\xi_{11}| \cos(\omega t + \arg \xi_{11}) \\ \dots \\ |\xi_{43}| \cos(\omega t + \arg \xi_{43}) \end{bmatrix} \quad (8)$$

## 6. Results of Computations

The mode shapes obtained from Eq. (8) with  $t = 0$  are shown with corresponding natural frequencies in Fig. 8 for the cases of 0- and 0.05-Hz constant spin. If the motion given by Eq. (8) is in the  $x_1x_2$  plane, arrows indicate the general direction of subsequent motion. The

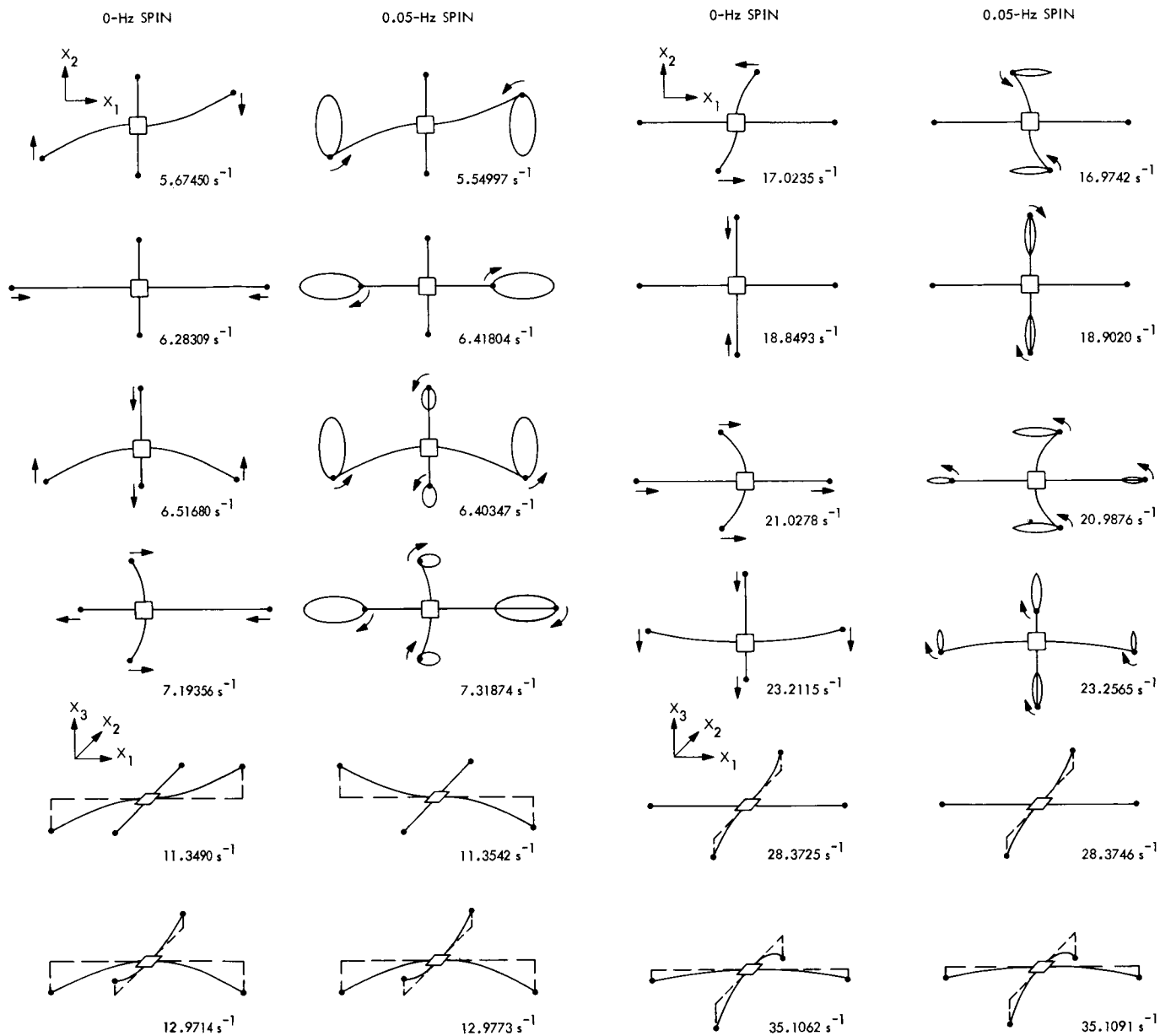


Fig. 8. Mode shapes and natural frequencies

paths described by the end masses are shown if they are elliptical; otherwise, they are straight-line segments or points.

The elliptical paths are attributable to the Coriolis effects represented by the  $G'$  matrix. Those modes having motion along the  $x_3$  coordinate line have velocities parallel to the spin vector; the Coriolis force is zero, and the mode shapes do not change significantly as spin is applied. The natural frequencies of these modes increase with increasing spin.

With no spin, the first mode involves pure bending, but as spin is invoked, the motion involves both bending and extension, and the natural frequency decreases. This decrease may seem unusual in view of the centrifugal forces of the steady motion. However, since the first two modes do not involve translation of the beam intersection, they may be obtained in the closed-form solution for a cantilevered beam rotating about a fixed axis through the built-in end. Frequencies obtained from this solution agreed exactly with those shown in Fig. 8. Had the extensional stiffness been much larger, say thirty times the bending stiffness, then both the frequencies of the first two modes would have increased with increasing spin.

### References

1. Likins, P. W., *Dynamics and Control of Flexible Space Vehicles*, Technical Report 32-1329. Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1969.
2. Bamford, R. M., *Application of Structural Analysis and Matrix Interpretive System*, Technical Memorandum 33-399. Jet Propulsion Laboratory, Pasadena, Calif., Oct. 15, 1968.

## E. Vibration and Deformation Measurements

**Using Holography Techniques, E. Heer, G. Morse, and R. Badin**

### 1. Introduction

The rapid development of holography during the last decade has brought this new technique within the range of practical application as a tool in various scientific and engineering disciplines. However, holography has not yet matured as a measuring tool to the point where standard instrumentation can be developed to measure, for example, mechanical vibrations and deformations. Such instrumentation is still in a state of trial-and-error development through applications to specific problem classes. The objective of this work is to develop applications

technology for the solution of problems using laser holographic techniques, and to develop breadboard instrumentation for the application to engineering mechanics problems. Areas presently under consideration are flaw detections, vibrations, and deformations.

### 2. Laboratory Installation

An existing laboratory was refurbished to make it suitable for exposing, developing, and viewing holograms. A stable platform was designed and constructed. Gaertner holographic equipment, consisting of a rectangular optical bench, a 3-mW helium-neon gas laser at 6328 Å wavelength, and assorted standard mirrors, plate holders, beam splitters, beam spreader, etc., was acquired and installed. In addition, a Spectral Physics 5-mW helium-neon gas laser and equipment for vibration and strain measurements by time averaging and by real-time holographic methods, were procured and installed for use. Figure 9 shows the Gaertner equipment set up with a cylindrical shell in the foreground ready for taking holograms.

### 3. Holography

The principles of holography have been published by many authors, as shown e.g. in Refs. 1 and 2. Considerable efforts have been spent during the last several months to develop holographic measuring techniques in three areas: (1) ordinary holography, (2) static deformation measurements, and (3) time-averaged vibration measurements.

An example of ordinary holography is shown in Fig. 10. The object is a human skull which can be seen in the

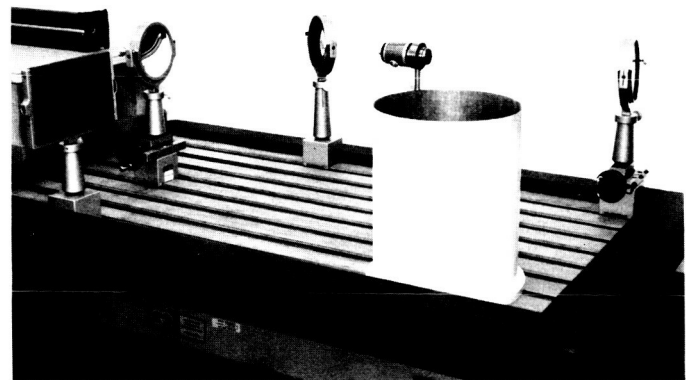


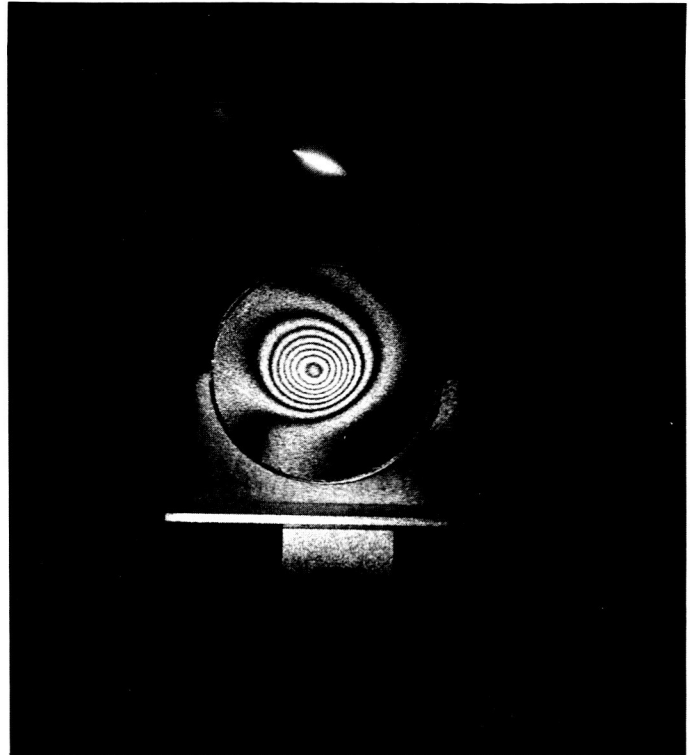
Fig. 9. Gaertner holography equipment



**Fig. 10. Photograph of a hologram of a human skull**



**Fig. 11. Solid propellant with pressurized internal cavity for pressure differential  $\Delta P_1$**

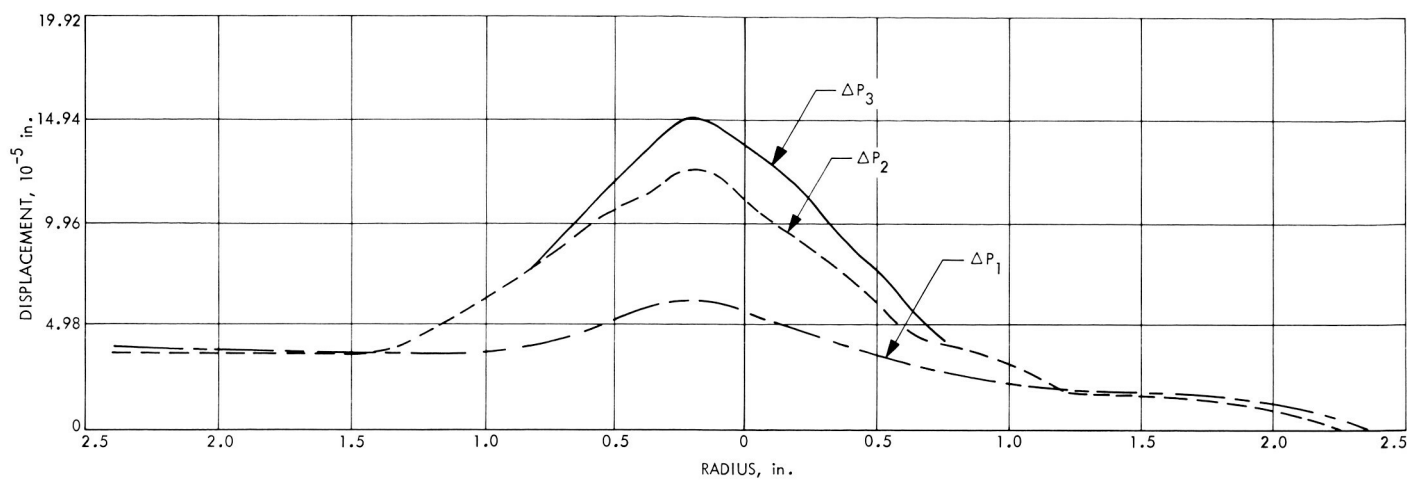


**Fig. 12. Solid propellant with pressurized internal cavity for pressure differential  $\Delta P_3$**

usual three-dimensional fashion when viewed in the laboratory (the photograph, of course, is two-dimensional). All holograms are presently made using Kodak 649F, or Agfa 10E75 4 × 5-in. plates, and Kodak D19 developer.

Figures 11 and 12 show a solid-propellant, 2-in.-thick disk with an internally located small cavity. A thin pipe connected to a pressure bottle leads to the cavity and allows the pressure within the cavity to vary. The fringe patterns in Figs. 11 and 12 were obtained by first exposing the hologram partially to one pressure level, and then performing a second exposure with a pressure increase of  $\Delta P_1$  and  $\Delta P_3$ , respectively. The location of the cavity is easily determined, and the amount of surface deformation because of the pressure changes can be obtained by counting the fringes. Since the wavelength  $\lambda$  is 6328 Å, the relative deformation between the centers of adjacent dark and bright areas is 1582 Å; i.e.,  $\frac{1}{4}$  of a wavelength. Figure 13 shows the surface deformations of the disk for three different  $\Delta P$  as a function of radial position.

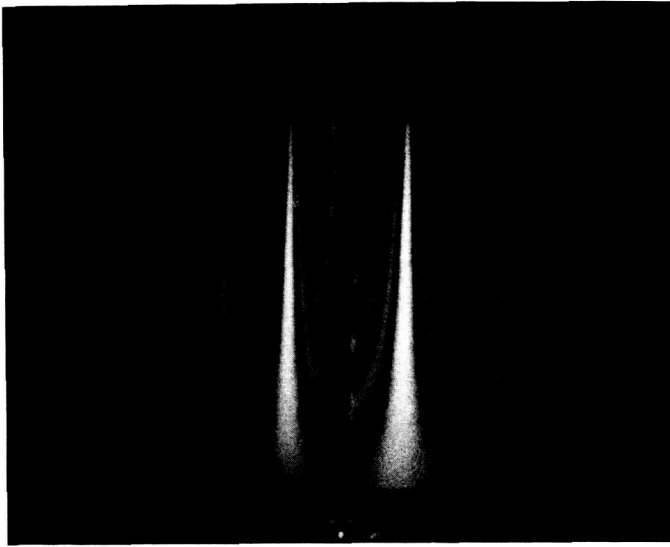
Figure 14 is a time-averaged hologram showing deformations of a vibrating circular plate. The plate was



**Fig. 13. Surface deformations of solid propellant with pressurized internal cavity**



**Fig. 14. Time-averaged hologram of a circular plate**

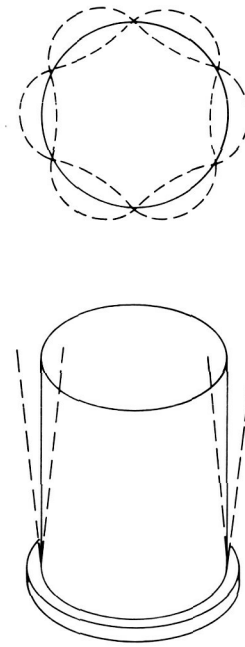


**Fig. 15. Time-averaged hologram of a cylindrical shell**

excited at the center point at 600 Hz while the complete exposure was obtained. Thus, each point on the plate surface gives an indication of the average energy intensity due to constructive or destructive interference for bright or light areas, respectively. A similar time-averaged hologram showing deformations of a vibrating circular cylindrical shell is shown in Fig. 15. The shell was excited at a calculated modal frequency of 150 Hz, which corresponds to an analytical modal deformation shown schematically in Fig. 16. It can be seen that the agreement of the analysis with the holographic presentation is very good.

#### 4. Conclusions

The application of holography to engineering mechanics problems in a routine fashion can be accomplished but needs additional work. In particular, the quantitative



**Fig. 16. Modal deformation of a cylindrical shell**

interpretation of measured results needs intensive consideration. As a next step in this work, additional measurements will be made of statically and/or dynamically deforming structural elements that can be easily analyzed using analytical techniques. This information will then be used to develop techniques for the quantitative interpretation of measured results.

#### References

1. Stroke, G. W., *An Introduction to Coherent Optics and Holography*, Second Edition. Academic Press, Inc., New York, 1969.
2. De Velis, T. R., and Reynolds, G. O., *Theory and Application of Holography*. Addison-Wesley Co., Reading, Mass., 1967.

## XII. Solid Propellant Engineering

### PROPULSION DIVISION

#### **A. Heat Transfer in a Burning Solid Propellant With Oscillating Surface Temperature and Surface Regression, R. L. Klaus**

##### **1. Introduction**

To properly describe solid propellant combustion, it is necessary to describe the heat transfer from the solid surface to the interior of the grain. This is accomplished through solutions of the heat conduction equation with boundary conditions chosen that correspond to the burning conditions being described. In transient combustion, both the boundary conditions and the motion of the surface vary with time. A general treatment of this problem has been given in SPS 37-60, Vol. III, pp. 187-199. One important specific set of boundary conditions is that which corresponds to oscillatory burning of the propellant. Both the surface temperature and motion of the surface are taken to oscillate about mean values.

If the heat conduction equation can be solved under these boundary conditions, it is possible to calculate the entire temperature profile in the solid. However, the value

of the derivative of the temperature at the surface, as a function of time, is of the greatest interest in modeling the overall combustion process. The knowledge of this derivative gives a means of calculating the rate of heat transfer into the solid at the surface, presuming knowledge of the instantaneous surface temperature and regression rate.

Solutions to the heat conduction equation are available for small oscillations in surface temperature and burning rate (Refs. 1 and 2). Because the oscillations are small, the equations can be linearized and easily solved. It is possible to use these results in conjunction with a gas-phase model to determine whether or not a given frequency of oscillation increases or decreases with time, and thus whether oscillatory burning is possible under those conditions. However, linearized theory, by its very nature, can never predict the amplitude of oscillations to be expected.

It is a well-known fact that under both laboratory and actual motor conditions, the amplitude of the oscillations in unstable burning is frequently not small compared to the mean values. This fact, in addition to the prospect

of eventually being able to predict the amplitudes of the oscillations, has led to the efforts discussed here; namely, more exact solutions for the heat conduction equation under conditions of oscillations of finite amplitude.

## 2. Mathematical Development

The mathematical treatment of this problem begins with the heat conduction equation for a solid with a moving surface:

$$\frac{\partial^2 T}{\partial x^2} + \left(\frac{r}{\alpha}\right) \frac{\partial T}{\partial x} = \frac{1}{\alpha} \frac{\partial T}{\partial t} - G(x, t) \quad (1)$$

where  $T$  is temperature,  $r$  is instantaneous regression rate of the solid or instantaneous velocity of the surface,  $\alpha$  is the thermal diffusivity,  $x$  is distance measured into the propellant,  $t$  is time and  $G(x)$  is a term that can account either for subsurface chemical reactions or radiation penetration. The surface temperature, motion of the surface and  $G$  are all taken to be periodic. This can be expressed as follows:

$$r = \sum_{n=0}^{\infty} \lambda^n r_n \exp [i(n\omega t)] \quad (2)$$

$$T(0, t) = \sum_{n=0}^{\infty} \lambda^n T_{0n} \exp [i(n\omega t)] \quad (3)$$

$$G(x, t) = \sum_{n=0}^{\infty} \lambda^n G_n(x) \exp [i(n\omega t)] \quad (4)$$

where  $\omega$  is a frequency,  $r_n$  and  $T_{0n}$  are complex constants,  $G_n(x)$  is a complex function of  $x$ ,  $\lambda$  is a small but finite parameter. By allowing  $r_n$  and  $T_{0n}$  to be complex, provision is made for the surface temperature and surface motion as well as their harmonics to be out of phase. The amplitude and phase angle of the  $n$ th harmonic of  $r$  are given by:

$$Z_n = (r_n^{(r)^2} + r_n^{(i)^2})^{1/2}$$

$$\phi_n = \arctan \frac{r_n^{(i)}}{r_n^{(r)}}$$

where  $r_n^{(r)}$  and  $r_n^{(i)}$  are the real and imaginary parts of  $r_n$ . Similar relationships apply for  $T_{0n}^*$ . The first terms in the summations are the nonperiodic terms. In the linearized theory, the first two terms are retained and the surface motion and surface temperature are considered to be exactly sinusoidal. For nonlinear analysis, additional terms must be retained.

The other two boundary conditions to the problem are:

$$T(\infty, t) = 0$$

$$T\left(x, t + \frac{2\pi}{\omega}\right) = T(x, t)$$

The first of these equations states that all temperatures are measured above the ambient temperature and that at distances sufficiently far into the solid, the ambient temperature prevails. The second of these equations states that the temperature at every point in the solid (not only the surface) is periodic in time but makes no commitment as to its phase relationship with the surface temperature.

The problem can be nondimensionalized to obtain

$$\frac{\partial^2 T^*}{\partial x^*} + r^*(t^*) \frac{\partial T^*}{\partial x^*} = \frac{\partial T^*}{\partial t^*} - G^*(x^*, t^*) \quad (5)$$

$$r^*(t^*) = \sum_{n=0}^{\infty} \lambda^n r_n^* \exp [i(nt^*)] \quad (6)$$

$$G^*(x^*, t^*) = \sum_{n=0}^{\infty} \lambda^n G_n^*(x^*) \exp [i(nt^*)] \quad (7)$$

with boundary conditions

$$T^*(0, t^*) = \sum_{n=0}^{\infty} \lambda^n T_{0n}^* \exp [i(nt^*)] \quad (8a)$$

$$T^*(\infty, t^*) = 0 \quad (8b)$$

$$T^*(x^*, t^* + 2\pi) = T^*(x^*, t^*) \quad (8c)$$

and where  $T_{00}^* = 1$ .

The nondimensionalization is

$$t^* = \omega t \quad x^* = \left(\frac{\omega}{\alpha}\right)^{1/2} x$$

$$T_{0n}^* = \frac{T_{0n}}{T_{00}} \quad T^* = \frac{T}{T_{00}}$$

$$r_n^* = \frac{r_n}{(\omega\alpha)^{1/2}}$$

$$G_n^*(x^*) = \left(\frac{\alpha}{\omega T_{00}}\right) G_n \left[\left(\frac{\omega}{\alpha}\right)^{1/2} x\right]$$

The definitions for  $r^*$  and  $G^*$  can now be substituted into the differential equation. In addition a solution of the form

$$T^*(x^*, t^*) = \sum_{n=0}^{\infty} \lambda^n T_n^*(x^*) \exp[i(n t^*)] \quad (9)$$

is assumed. When these relationships are substituted into the original equation, and like powers of the parameter  $\lambda$  equated,

$$\frac{d^2 T_n^*}{dx^{*2}} + r_0^* \frac{dT_n^*}{dx^*} - in T_n^* = V_n(x^*) \quad (n = 0, 1, 2, \dots, n) \quad (10)$$

is obtained where

$$V_n(x^*) = \sum_{j=1}^n r_j^* \frac{dT_{n-j}^*}{dx^*} - G_n^*(x^*) \quad (n = 1, 2, \dots, n) \quad (11a)$$

$$V_0(x^*) = G_0^*(x^*) \quad (11b)$$

with boundary conditions

$$T_n^*(0) = T_{n0}^* \quad T_n^*(\alpha) = 0$$

The partial differential equation has been reduced to an ordinary differential equation. Also, the third boundary condition has been satisfied by the choice of the form of the solution,  $T^*(x^*, t^*)$ . Similar equations can be written for all  $n$  of interest. Equations (11a) and (11b) are really recursion relationships in that  $V_n$  is expressed in terms of  $dT_j^*/dx^*$  for  $j < n$ .

Equation (10), and the boundary conditions, have the following solution:

$$\begin{aligned} T_n^* = & T_{n0}^* \exp[-(k_1 + k_2^{(n)})x^*] \\ & + \frac{1}{2k_2^{(n)}} \left\{ \int_0^\infty V_n(\xi) \exp[k_1(\xi - x^*) - k_2^{(n)}(\xi + x^*)] d\xi \right. \\ & - \int_0^{x^*} V_n(\xi) \exp[-(k_1 + k_2^{(n)})(x^* - \xi)] d\xi \\ & \left. - \int_{x^*}^\infty V_n(\xi) \exp[-(k_1 - k_2^{(n)})(x^* - \xi)] d\xi \right\} \end{aligned}$$

where

$$k_1 = \frac{r_0^*}{2} \quad k_2^{(n)} = \left[ \left( \frac{r_0^*}{2} \right)^2 + in \right]^{1/2} \quad (12)$$

The derivative of the temperature can be calculated to be

$$\begin{aligned} \frac{dT_n^*}{dx^*} = & -[k_1 + k_2^{(n)}] T_{n0}^* \exp[-(k_1 + k_2^{(n)})x^*] - \left[ \frac{k_1 + k_2^{(n)}}{2k_2^{(n)}} \right] \left\{ \int_0^\infty V_n(\xi) \exp[k_1(\xi - x^*) - k_2^{(n)}(\xi + x^*)] d\xi \right. \\ & \left. - \int_0^{x^*} V_n(\xi) \exp[-(k_1 + k_2^{(n)})(x^* - \xi)] d\xi \right\} + \left[ \frac{k_1 - k_2^{(n)}}{2k_2^{(n)}} \right] \int_{x^*}^\infty V_n(\xi) \exp[-(k_1 - k_2^{(n)})(x^* - \xi)] d\xi \end{aligned}$$

To reduce the complexities of these equations, it is assumed that  $V_n$  has the following form:

$$V_n = \sum_{m=0}^{n-1} A_m^{(n)} \exp(-B_m x^*) \quad (13)$$

When substituted into the equations for  $T_n^*$  and  $dT_n^*/dx^*$  and the indicated integrations are performed, the following are obtained:

$$T_n^* = \exp[-(k_1 + k_2^{(n)})x^*] \times \left\{ T_{n0}^* - \sum_{m=0}^{n-1} \frac{A_m^{(n)}}{[(k_1 - B_m)^2 - k_2^{(n)2}]} \right\} + \sum_{m=0}^{n-1} \frac{A_m^{(n)} \exp(-B_m x^*)}{[(k_1 - B_m)^2 - k_2^{(n)2}]}$$

and

$$\frac{dT_n^*}{dx^*} = -(k_1 + k_2^{(n)}) \exp[-(k_1 + k_2^{(n)})x^*] \times \left\{ T_{n0}^* - \sum_{m=0}^{n-1} \frac{A_m^{(n)}}{[(k_1 - B_m)^2 - k_2^{(n)2}]} \right\} - \sum_{m=0}^{n-1} \frac{A_m^{(n)} B_m \exp(-B_m x^*)}{[(k_1 - B_m)^2 - k_2^{(n)2}]}$$

If  $B_m$  is taken to be

$$B_m = k_1 + k_2^{(m)} \quad (14)$$

then  $dT_n^*/dx^*$  can be expressed in the following compact form:

$$\frac{dT_n^*}{dx^*} = \sum_{m=0}^n C_m^{(n)} \exp(-B_m x^*) \quad (15)$$

$$C_m^{(n)} = \frac{-i A_m^{(n)} B_m}{(n-m)} \quad m < n \quad (16a)$$

$$C_n^{(n)} = -B_n \left[ T_{n0}^* - \sum_{m=0}^{n-1} \frac{i A_m^{(n)}}{(n-m)} \right] \quad n > 0 \quad (16b)$$

$$C_0^{(0)} = -B_0 T_{00}^* \quad (16c)$$

These equations necessitate a knowledge of  $A_m^{(n)}$  that can be obtained from a comparison of Eqs. (11a), (11b), (13), and (15) as follows:

$$\left. \begin{aligned} \sum_{m=0}^{n-1} A_m^{(n)} e^{-B_m x^*} &= \sum_{j=0}^{n-1} r_{n-j}^* \sum_{m=0}^j C_m^{(j)} e^{-B_m x^*} \\ A_m^{(n)} &= \sum_{j=m}^{n-1} r_{n-j}^* C_m^{(j)} \end{aligned} \right\} \quad (17)$$

With Eqs. (14), (15), (16a), (16b), and (16c), this provides a means for the calculation of  $dT_n^*/dx_n^*$ . Equation (17) also verifies that the forms chosen for  $V_n$  and  $B_n$  in Eqs. (13) and (14) were correct. Since Eq. (9) may be differentiated directly, it is possible to calculate

$$\frac{\partial T^*}{\partial x^*}(x^*, t^*) = \sum_{n=0}^{\infty} \lambda^n \frac{dT_n^*}{dx^*}(x^*) \exp[i(n t^*)] \quad (18)$$

which is the derivative of greatest interest. (For the sake of simplicity, the term  $G_n^*$  in Eq. 11a has been ignored. However, a treatment which includes them follows the same principles.) It is usually of interest to evaluate it at the surface, which corresponds to  $x^* = 0$ . This calculation depends on a knowledge of the oscillatory motion of the burning surface and surface temperature which enter through  $r_n^*$  and  $T_{n0}^*$ . Equation (15) is perfectly general in that it is capable of accounting for oscillations and harmonics of finite amplitude and phase angle. Practically, it is important that the amplitudes be sufficiently small so that an undue number of terms need not be calculated in Eq. (18). This is not a serious consideration for oscillations of practical interest.

The temperature profile throughout the solid may be calculated by

$$T_n^*(x^*) = \sum_{m=0}^n D_m^{(n)} e^{-B_m x^*}$$

$$D_m^{(n)} = \frac{i A_m^{(n)}}{(n-m)} \quad m < n$$

$$D_n^{(n)} = T_{n0}^* - \sum_{m=0}^{n-1} \frac{i A_m^{(n)}}{(n-m)} \quad n > 0$$

$$D_0^{(0)} = T_{00}^*$$

### 3. Application to Oscillatory Combustion

A computer program has been written to perform the calculations discussed in Subsection 2. The instantaneous motion of the surface was taken to be directly related to the instantaneous surface temperature according to the following pyrolysis law:

$$r = A \exp(-B/T_s) \quad (19)$$

where  $A$  and  $B$  are empirical constants.

Because Eq. (2) demands that the motion of the surface be expressed in the same way as the surface temperature, some manipulation of Eq. (19) is necessary. The simplest way to express  $r$  in suitable form is through a Fourier series expansion. The functions  $[\exp(imx)]$  are orthogonal over the interval  $0 \leq x \leq 2\pi$ ,

$$\int_0^{2\pi} \exp(imx) \exp(inx) dx = \begin{cases} 0 & m \neq -n \\ 2\pi & m = -n \end{cases}$$

Both  $r_n^*$  and  $T_{n0}^*$  may be written in their expanded form:

$$\sum_{m=0}^{\infty} \lambda^m r_m^* \exp(imt^*) = A \exp \left[ \sum_{m=0}^{\infty} \lambda^m T_{0m}^* \exp(imt^*) \right]$$

If this equation is multiplied through by  $\exp(inx)$  and integrated over the interval  $0 \leq x \leq 2\pi$ , all the terms but one drop out of the left side of the equation and one obtains an explicit expression for  $r_n^*$ ,

$$\lambda^n r_n^* = A \int_0^1 \exp \left[ \frac{-B}{\sum_{m=0}^{\infty} \lambda^m T_{0m}^* \exp(2\pi i m \tau) - 2\pi i n \tau} \right] d\tau \quad (20)$$

which is in the proper form for use in the calculation of the surface derivative of temperature.

#### 4. Discussion

The computer program has provision to read in  $T_{0n}^*$  as a set of complex numbers. The correspondence between the real and imaginary parts of  $T_{0n}^*$  and the amplitude and phase angle of the oscillation harmonics is the same as for  $r_n^*$  discussed previously. There are two options in the program with regard to  $r_n^*$ . Either  $A$  and  $B$  may be read in and  $r_n^*$  calculated according to Eq. (20) or  $r_0^*$  and  $r_1^*$  are read in. In the latter case, values of  $A$  and  $B$  that correspond to the given values are  $r_0^*$  and  $r_1^*$  are calculated and these constants are used to generate the rest of the  $r_n^*$ . A set of  $r_n^*$  is calculated successively until the most recent calculated value is negligible compared to  $r_0^*$ . Although the  $r_n^*$  may actually be complex numbers, for the case in which all  $T_{0n}^*$  are real (and hence in phase), all  $r_n^*$  will be real (and hence in phase, both with each other, and with the  $T_{0n}^*$ ).

The program calculates  $dT^*/dx^*$  at  $x^* = 0$  by calculating the set of constants,  $C_m^{(n)}$ . In addition to generating these constants, it has a provision for generating curves for surface temperature, surface motion, and surface temperature derivatives as functions of time so that the wave form may be inspected.

Some of the results obtained through the use of the computer program are shown in Figs. 1-4. In Figs. 1 and 2 the amplitude and phase angle of  $dT_{01}^*/dx^*$  are shown plotted against  $r_0^*$  with  $r_1^*/T_{01}^*$  as a parameter. By plotting the ratio of the amplitudes of  $dT_{01}^*/dx^*$  and  $T_{01}^*$ , a great deal of information may be plotted on one curve since this ratio is constant for given values of  $r_0^*$  and  $r_1^*/T_{01}^*$ . As shown in Fig. 1, the curves become linear at high values of  $r_0^*$ . Values of the amplitude ratio were also calculated for  $r_1^*/T_{01}^* < 1$ , but these tend to cluster very close to the curve for  $r_1^*/T_{01}^* = 0$  and were not plotted. The same was true for the phase angle curves shown in Fig. 2, though not to the same extent.

Because the rate of heat transfer is opposite in sign to  $dT^*/dx^*$ , the phase angle of  $-dT^*/dx^*$  is actually calculated and plotted. A positive phase angle implies that absolute value of the heat transfer rate leads the surface temperature and a negative value implies that it lags.

The results shown in Figs. 1 and 2 are identical to what would be obtained from linearized theory because they deal only with the fundamental oscillation. To illustrate the departure from linearity, the ratio of the amplitudes of  $T_{02}^*$  and  $T_{01}^*$  is plotted in Fig. 3. If non-

linearity were ignored, all these curves would be identically zero. It is clear that for low  $r_0^*$  and high  $r_1^*/T_{01}^*$ , the second harmonic (and presumably some of the higher ones) cannot be ignored.

In Fig. 4, the actual temperature derivative at the surface is plotted against  $t^*$  for one cycle for typical values of  $T_{01}^*$ ,  $r_0^*$ , and  $r_1^*$ . Both the linear approximation and the exact solution are shown. In order to obtain values of 1 and 0.3 for  $r_0^*$  and  $r_1^*$ , the computer calculated values of 2.71828 and 1 for  $A$  and  $B$ , respectively, for use in Eq. (19). It then calculated values of  $r_2^*$  through  $r_8^*$  by Eq. (20). They were all real and decreasing in magnitude. Higher ordered  $r_n^*$  were discarded as not being significant to the calculation. From the significant values of  $r_n^*$  and  $T_{01}^*$ , the temperature derivative at the surface was calculated and plotted on Fig. 4. The average value was subtracted out so that the derivative appears to oscillate about zero rather than about the average rate (equal to one in this case). The resultant curve for the first harmonic is the linear approximation and is, of course, exactly sinusoidal at a phase angle of 9.57 deg and amplitude of 0.586. The program calculated  $dT_n^*/dx^*$  at  $x^* = 0$  through  $n = 7$ , and judged higher ordered terms not to be significant. A curve based on these constants is also shown in Fig. 4. It departs slightly from sinusoidal shape and this departure would increase for higher values of  $T_{01}^*$  and  $r_1^*$ .

#### 5. Concluding Remarks

It has been shown that accurate values of the instantaneous heat transfer at the surface may be calculated when full account is taken of the nonlinearities introduced by the presence of finite oscillations in both surface temperature and surface regression rate. These nonlinearities produce small but detectable departures from sinusoidal heat transfer rate predicted by linearized theory. The calculations are not difficult, particularly with the aid of a computer. This approach now gives a means of calculating the exact heat transfer response in situations where these nonlinearities may be significant.

#### References

1. Price, E. W., and Culick, F. E. C., "Combustion of Solid Rocket Propellants," *AIAA Professional Study Series*, 1968.
2. Culick, F. E. C., "A Review of Calculations of the Admittance Function for a Burning Surface," Appendix A of *Combustion of Solid Propellants and Low Frequency Combustion Instability*, NOTS TP 4244, U.S. Naval Weapons Center, China Lake, Calif., June 1967.

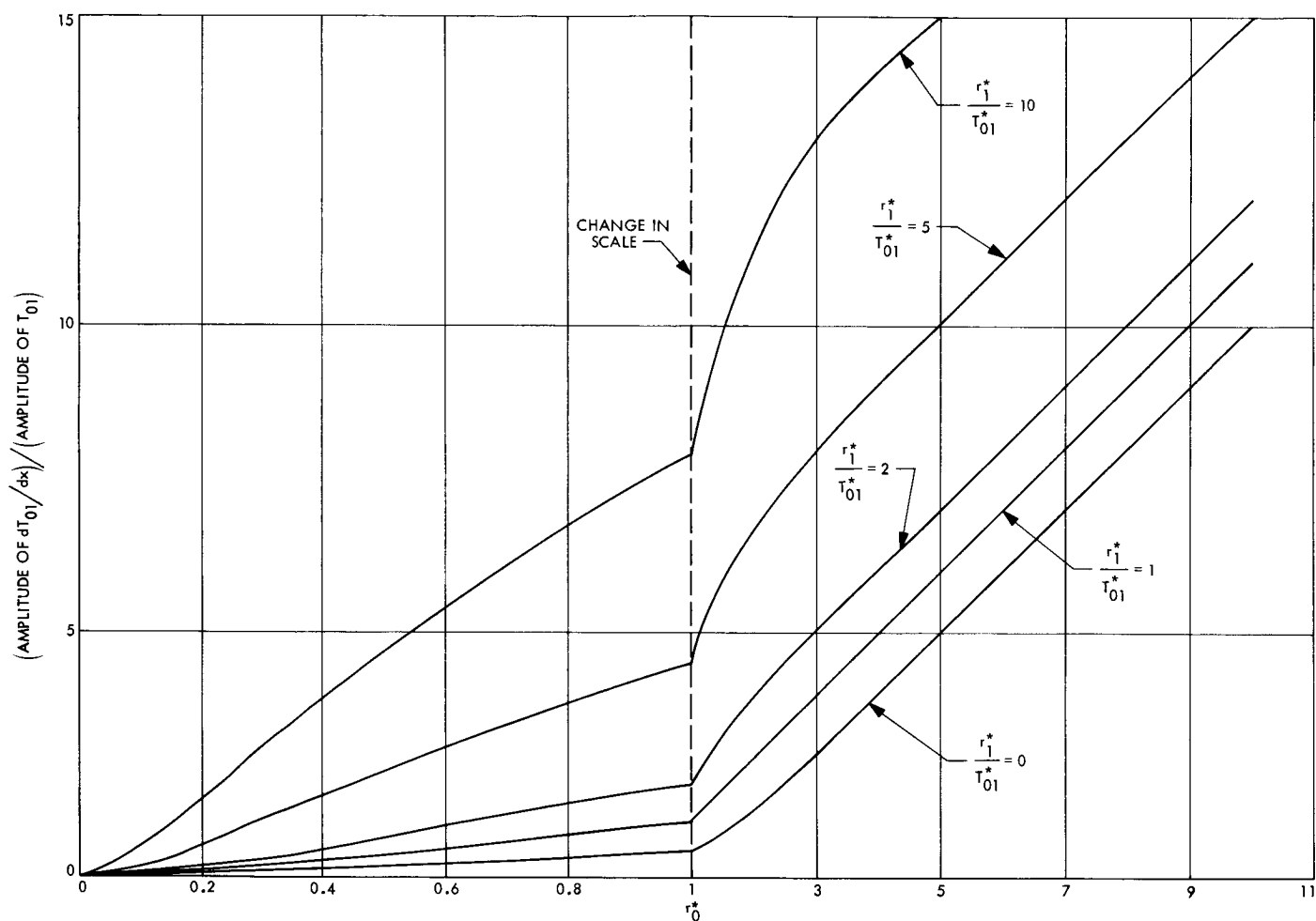


Fig. 1. Ratio of the amplitudes of  $dT_{01}^*/dx^*$  and  $T_{01}^*$  vs  $r_0^*$

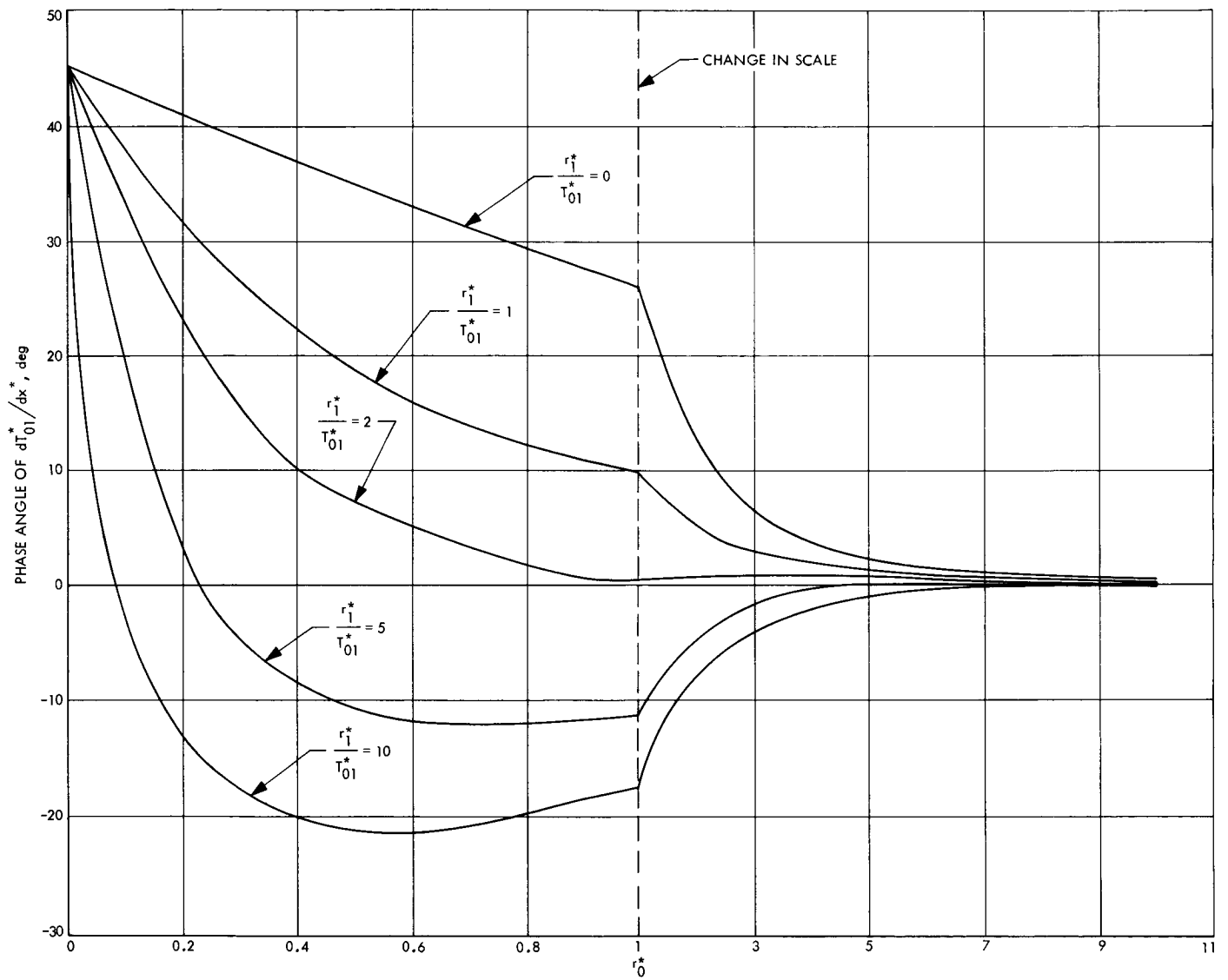


Fig. 2. Phase angle of  $dT_{01}^*/dx$  vs  $r_0^*$

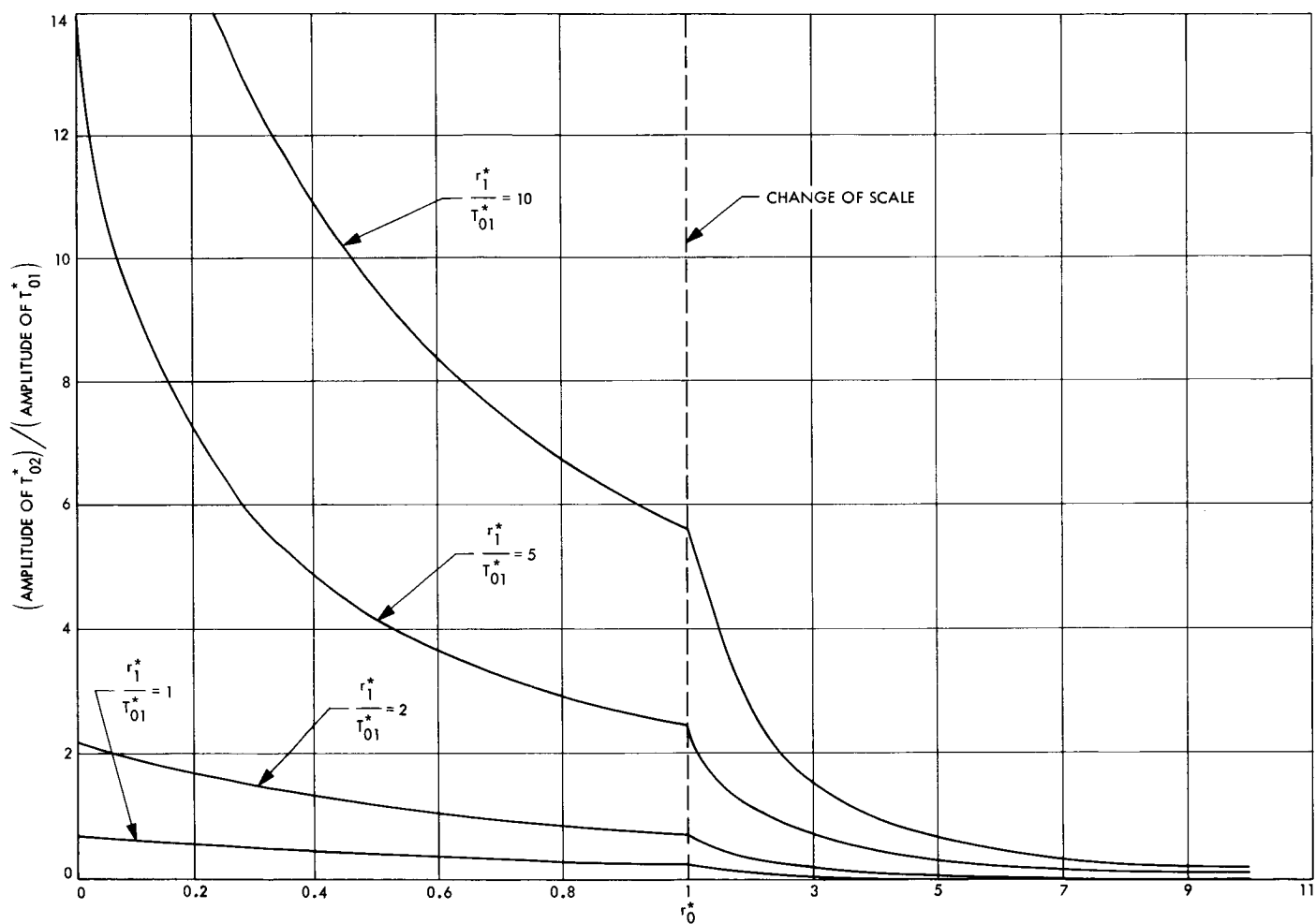
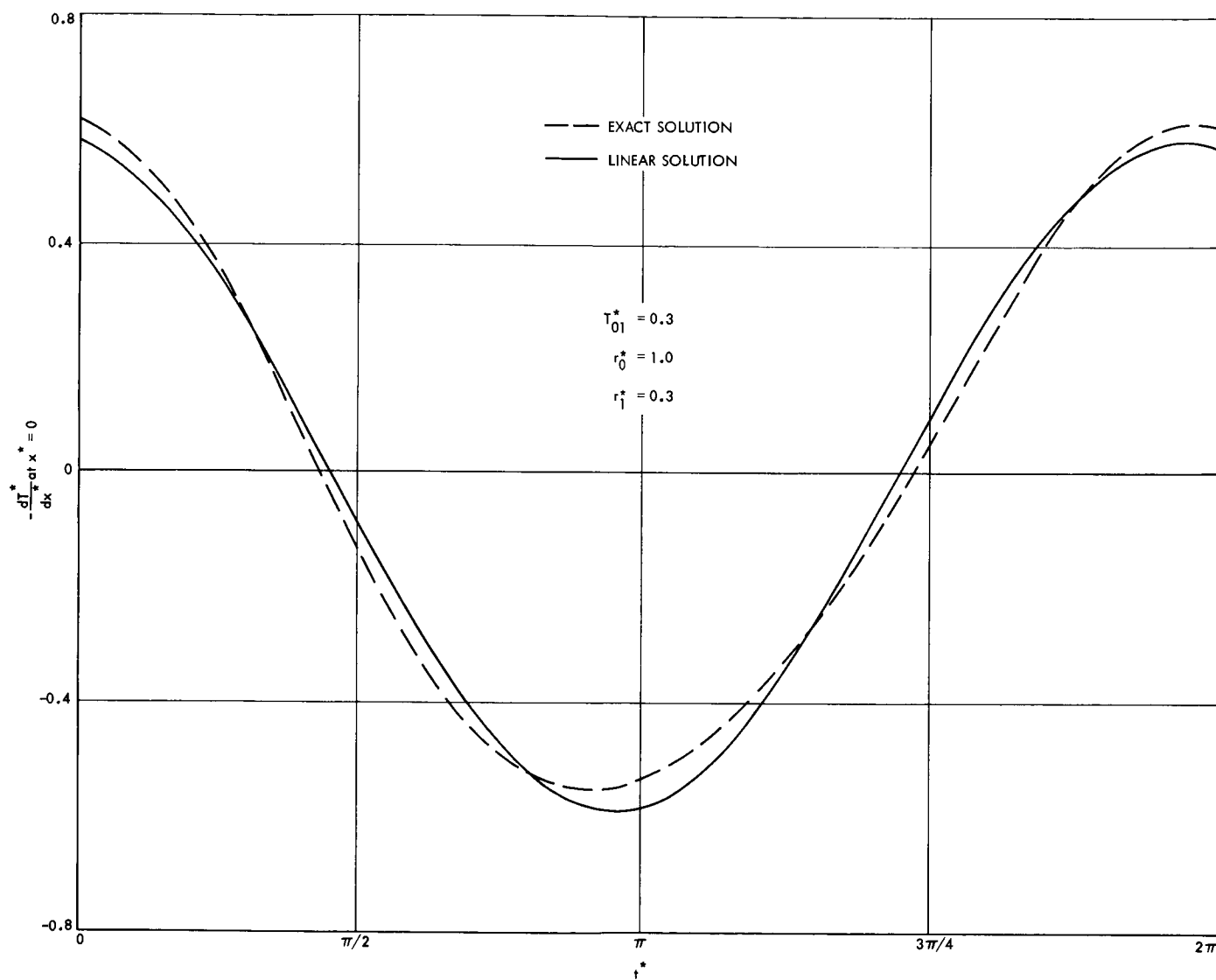


Fig. 3. Ratio of the amplitude of  $T_{02}^*$  and  $T_{01}^*$



**Fig. 4. Temperature derivative at the surface**

## **B. Experimental Study of the Combustion Termination of a Solid Propellant by Rapid Depressurization, W. Gerber**

### **1. Introduction**

Rapid depressurization of solid propellant rockets as a mechanism for terminating combustion has been studied by a number of investigators. Termination of combustion or "quench" is obtained by rapidly opening one or more secondary exhaust orifices on command, usually by puncturing or bursting some type of diaphragm. Trends can be observed from existing experimental data of the ease with which propellants of varying composition can be extinguished. However, knowledge of propellant composition, burning rate characteristics, and motor configuration, does not make it possible to quantitatively predict the minimum depressurization rate required to terminate thrust at any given motor chamber pressure.

A test program was initiated to gain information concerning the depressurization rates required to quench a new JPL solid propellant, called Saturethane, over the relatively low pressure interval of 65–200 psig. The propellant consists of a saturated polybutadiene binder, an ammonium perchlorate oxidizer, and an aluminum additive (16% by weight). For comparison, a few tests were also conducted with JPL 540 propellant, which has a polyurethane binder system.

### **2. Test System**

A primary guideline was to use as much existing hardware as possible. A window motor, described in SPS 37-52, Vol. III, pp. 89–96, was modified for use as a test device (Fig. 5). The motor nozzle was replaced with a Fike safety burst diaphragm assembly to which had been added a sized orifice. After ignition of the propellant, pressure continued to build rapidly in the closed motor until the diaphragm burst at the designed chamber pressure. The pressure decay rate was controlled by varying the diameter of the exit orifice (Fig. 6). The initial chamber pressure rise rate was adjusted by selecting the size of the propellant burning area. The tests were conducted with 1¼-in.<sup>2</sup> by ½-in.-thick slabs of propellant, restricted on the ½-in. sides with a ¼-in.-thick coating of vinyl cement. The slabs were ignited by a nichrome hot-wire pressed into slits cut in the propellant surface (Fig. 7). Rapid and uniform ignitions were attained in this manner.

Test data consisted of motor chamber pressure (Dynisco pressure transducer) and the combustion light

intensity, as determined by a silicon diode photovoltaic cell. The latter was used to indicate cessation of combustion.

### **3. Test Program**

Twenty-five depressurization tests were conducted with from two to seven firings at each of the following nominal chamber pressures: 65, 85, 120, 170, 200, and 265 psig. A reproduction of a typical test where quenching of combustion occurred is shown in Fig. 8. In a majority of the quench tests, the photocell trace showed a point of inflection followed by a steady decrease in output to the ambient level. A decrease in the pressure venting rate caused the inflection point to occur at a time corresponding to a lower chamber pressure. This inflection point was assumed to be the beginning of combustion termination. A typical no-quench test record is illustrated in Fig. 9. For all but the smallest venting orifice sizes, the small propellant burning area was unable to maintain sonic flow through the orifice, and the motor vented to ambient pressure. The photocell signal would continue at a decreased level until propellant burnout.

### **4. Test Results**

Other investigators have had difficulty correlating this type of test data with a simple set of parameters. They have concluded that the entire depressurization pressure-time history up to the onset of quench is important in determining the success or failure of quenching (Refs. 1 and 2). Nevertheless, it was deemed adequate for this study to plot the data as the initial depressurization rate vs the chamber pressure and the same two parameters measured at the onset of quench, as determined from the photocell records. The former method has been widely used by other investigators, and the latter are the correlation parameters predicted by an early theoretical model of the phenomenon (Ref. 3).

The test data are presented in Table 1 and are plotted in Figs. 10 and 11. A test was considered to be a quench if the photocell showed no deflection for several seconds following the depressurization. In many tests the propellant would reignite and burn to completion. Reignition is considered to be a separate problem from that of quench, and was not investigated here.

### **5. Discussion**

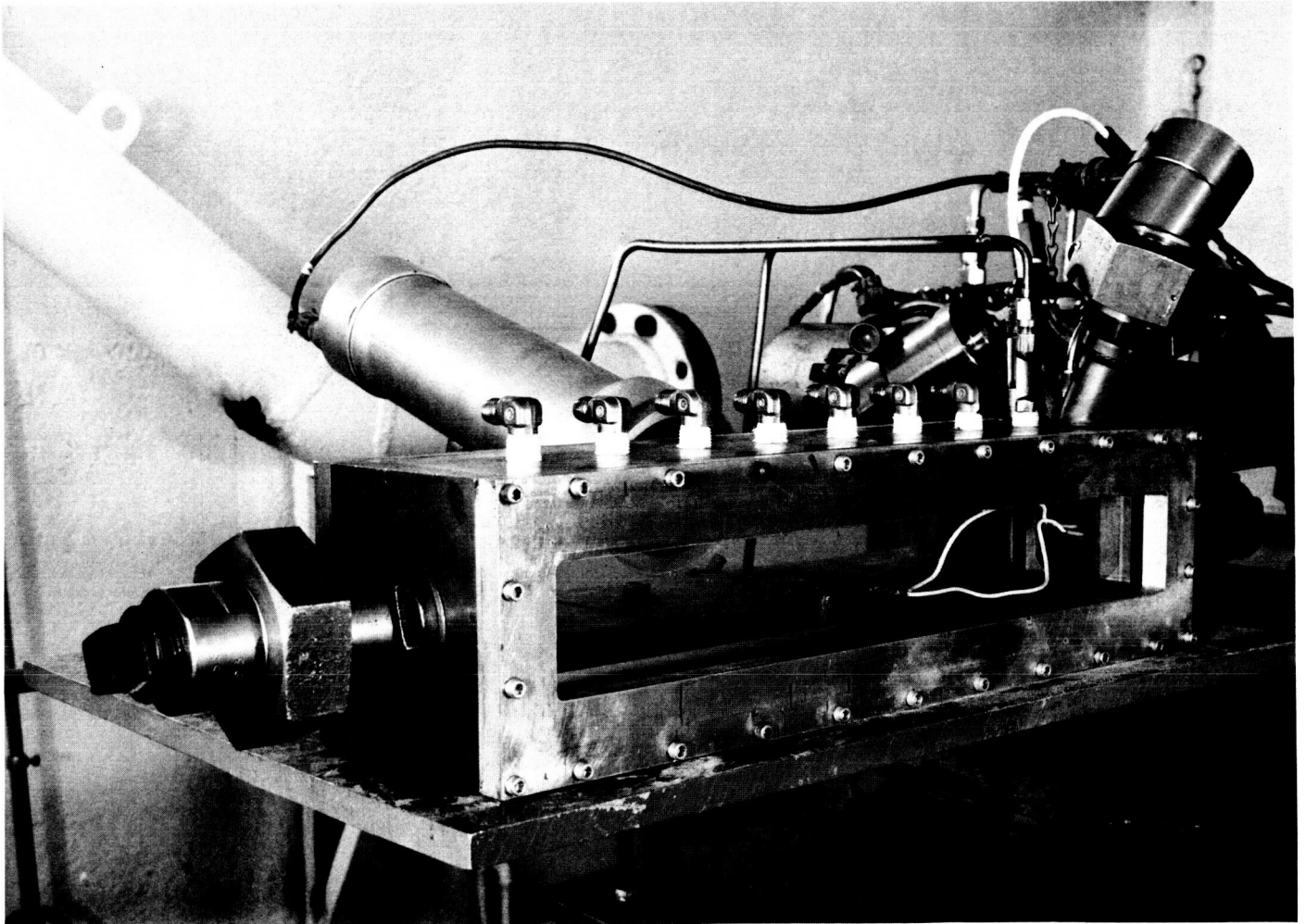
Discussion of the data will be centered around Fig. 10. The sequence of tests started at the lowest chamber

Table 1. Test data

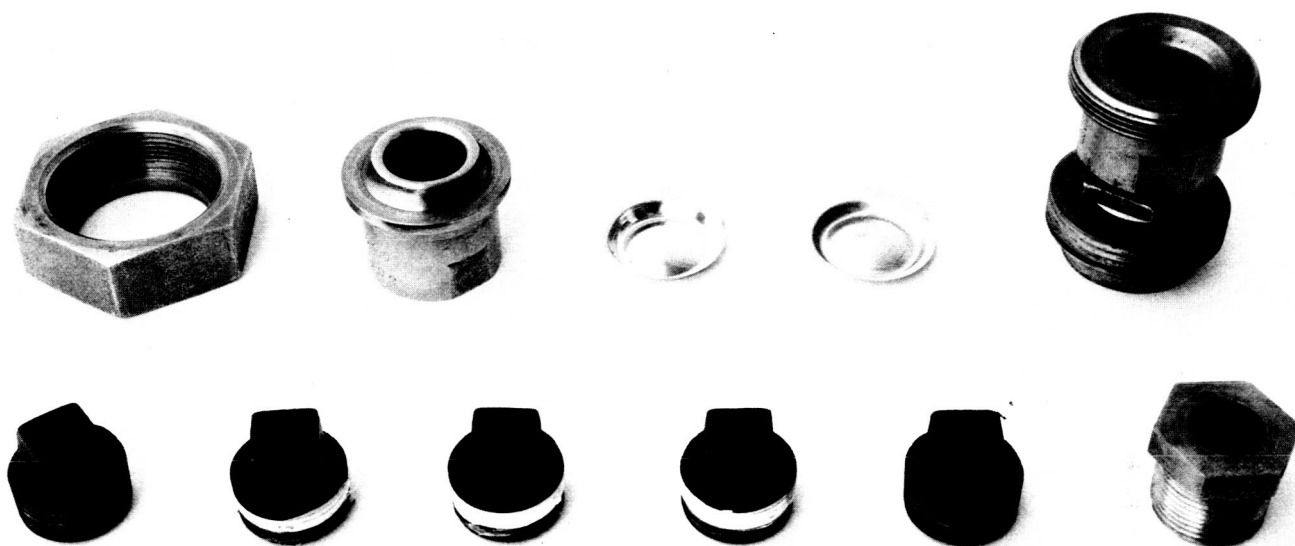
Test No.	Orifice diam, in.	Burst pressure, psig	Initial depressurization rate, psi/s	Pressure at quench, psig	Depressurization rate at quench, psi/s	Test results
1718 <sup>a</sup>	0.844	78	14,000	—	—	Quench
1718A	0.312	75	940	18	520	Quench
1718B	0.125	139	190			No quench
1718C	0.234	145	750	20	180	Quench
1718D	0.187	65	620	23	200	Quench
1718F	0.125	67	80			Marginal
1727	0.312	124	1,800	18	400	Quench, reignited
1727A	0.375	121	2,800	28	850	Quench
1727C	0.375	122	2,800	51	1,400	Quench
1727D	0.125	117	740			No quench
1728	0.375	211	4,700	6	630	Quench, delayed
1728A	0.437	199	6,200	42	1,300	Quench
1728B	0.234	196	1,600	11	200	Quench
1728C	0.187	192	1,500			No quench
1730	0.74	276	16,500			No quench partial
1731	0.44	263	7,500			No quench
1731A	1.00	263	32,000			No quench
1733	1.00	207	25,000			No quench
1734	1.00	194	23,000			Quench, delayed
1735	0.44	203	6,200			No quench
1747	0.84	176	15,000			No quench
1747A	1.00	170	—			No quench
1747B <sup>b</sup>	0.31	122	2,200	—	—	Quench
1747C	0.125	124	1,100			No quench
1750	0.187	117	1,300			No quench

<sup>a</sup>Tests 1718–1747A were conducted with Saturethane propellant.

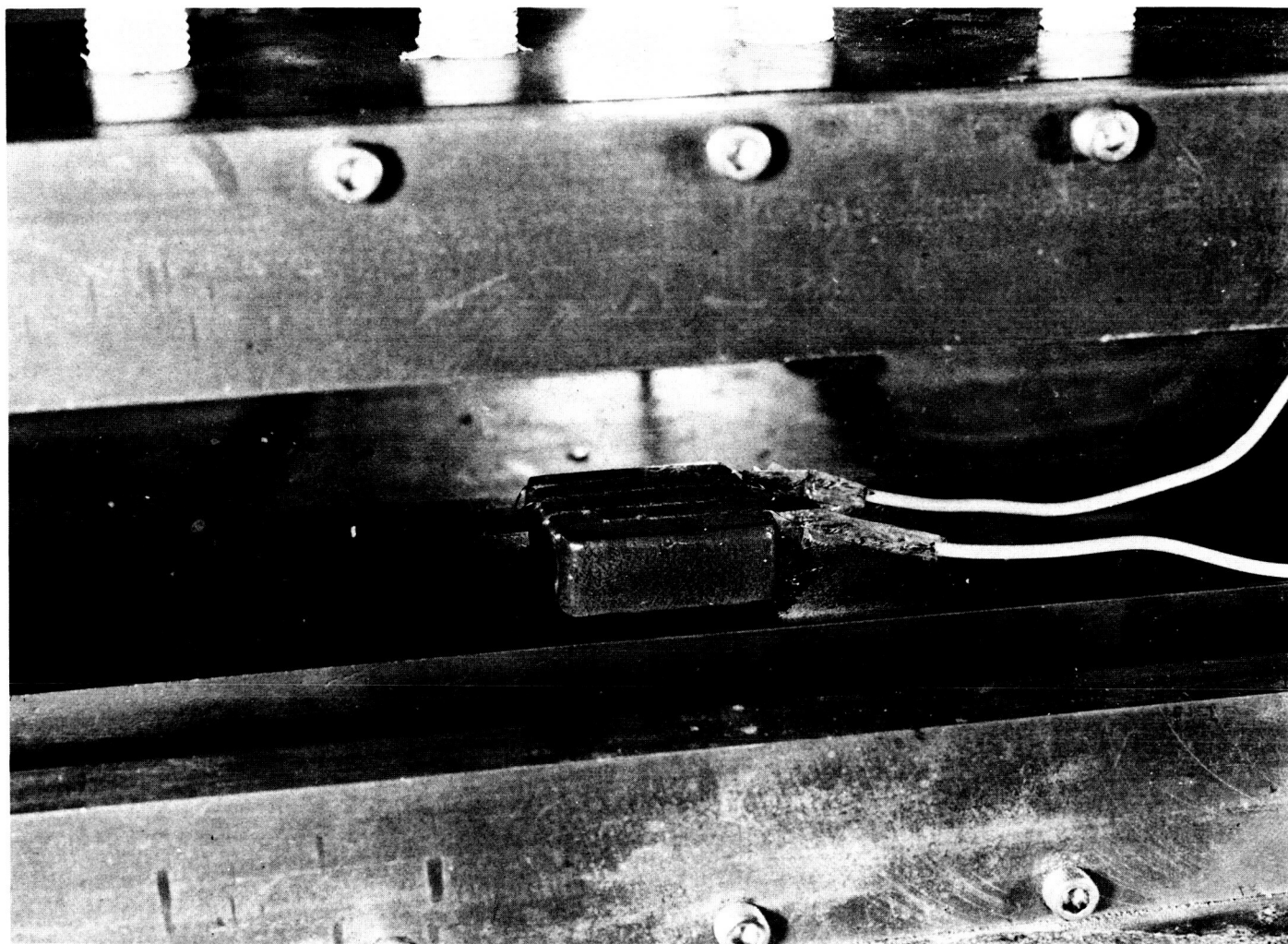
<sup>b</sup>Tests 1747B–1750 were conducted with JPL 540 propellant.



**Fig. 5. Test system**



**Fig. 6. Disassembled burst diaphragm holder and set of exhaust orifices**



**Fig. 7. Propellant slab**

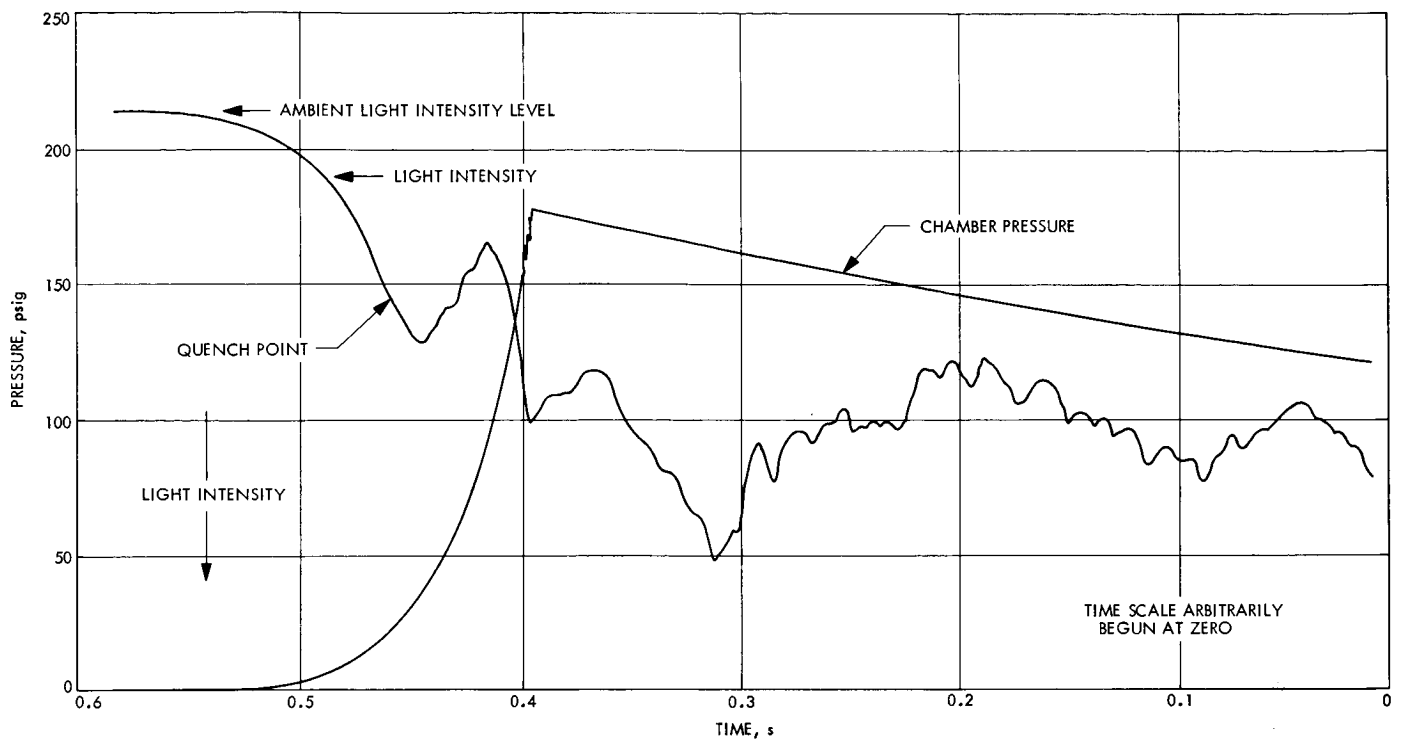


Fig. 8. Typical quench test record

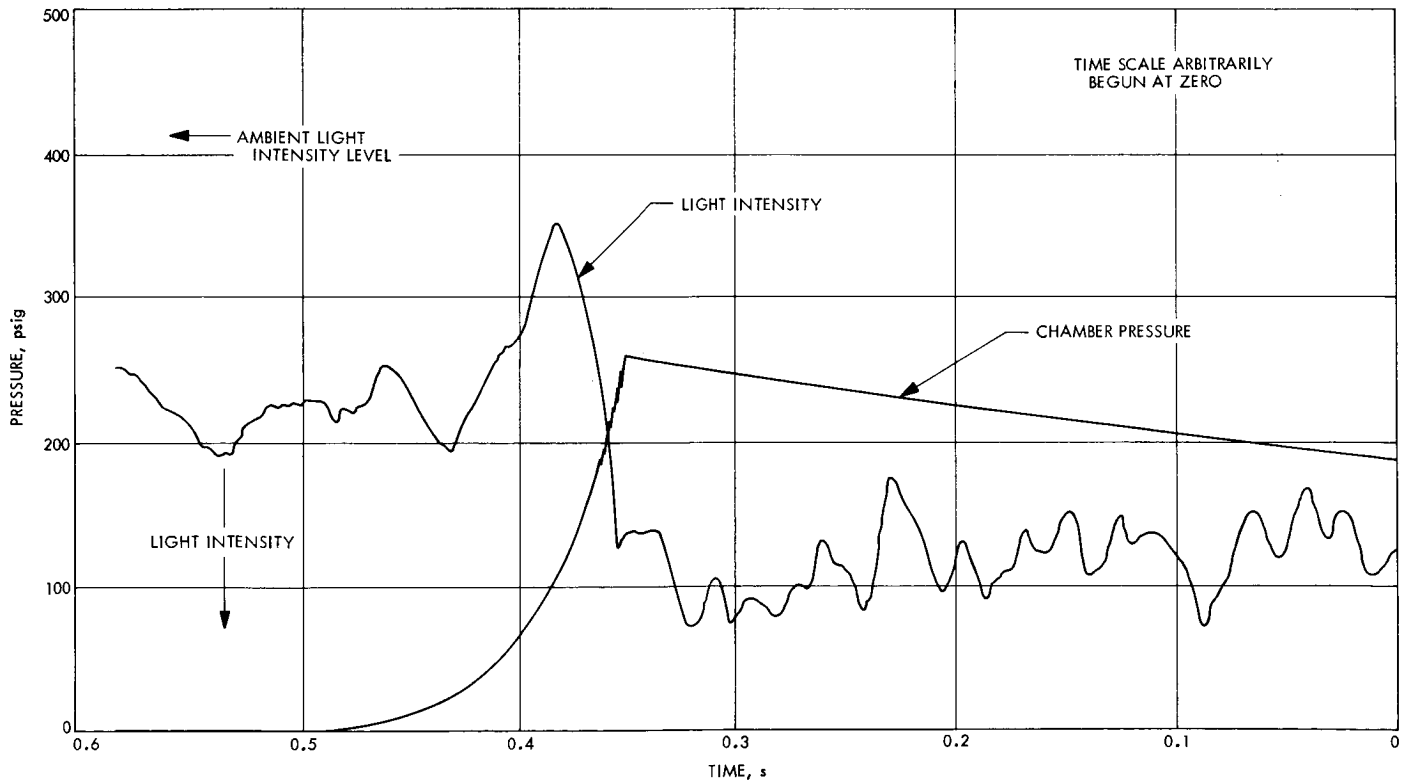


Fig. 9. Typical no-quench test record

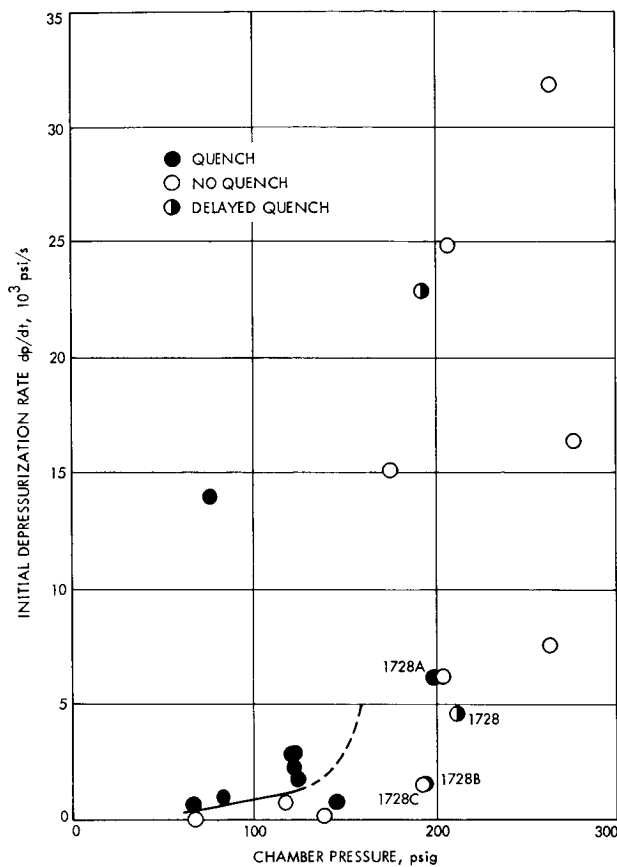


Fig. 10. Initial depressurization rate vs chamber pressure

pressure and continued up to the highest level. Quench was readily attained over the originally intended interval of 65–200 psig, so it was decided to extend the tests to higher pressures. Three tests at chamber pressure levels in the 265-psig range all yielded no quenches. The maximum depressurization rate, with the maximum orifice diameter of 1.0 in., was 32,000 psi/s. Additional tests were then run at 170 and 200 psig to verify that such an abrupt change in the quench requirements did actually exist. In five tests no quenches were obtained. Several additional tests conducted at 120 psig with JPL 540 propellant were in good agreement with the earlier Saturethane data. The quench point data (Fig. 11), were correlated reasonably well by the two parameters.

## 6. Conclusions

No theory has been derived to explain why quenches were initially obtained at a chamber pressure of 200 psig (runs 1728, 1728A, and 1728B) since later tests resulted in no quenches obtained in this motor at pressures greater than 170 psig. It is tentatively concluded that the quench depressurization requirements for the Sature-

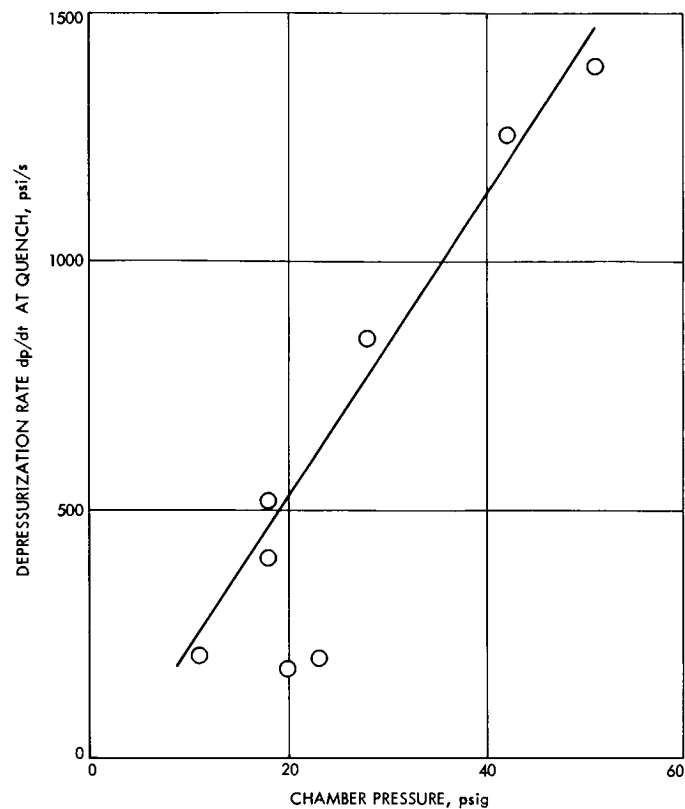


Fig. 11. Depressurization rate vs chamber pressure at quench point

thane propellant are reasonably moderate (initial rates less than 1000 psi/s) for pressures less than 100 psig, but increase rapidly for pressures greater than this value.

Additional tests at an intermediate pressure are planned in an attempt to define the speculative critical quench boundary region indicated by the dashed line in Fig. 10.

## References

1. Jenson, G. E., *Experimental Study of Solid Propellant Extinguishment by Rapid Depressurization*, Report R-UTC-2311-FR. Also available as NASA CR-66747, National Aeronautics and Space Administration, Washington, United Technology Center, Sunnyvale, Calif., Oct. 1970.
2. Merkle, C. L., and Turk, S. L., and Summerfield, M., "Extinguishment of Solid Propellants by Depressurization: Effects of Propellant Parameters," Paper 69-176, presented at the AIAA Seventh Aerospace Sciences Meeting, New York, Jan. 1969.
3. von Elbe, G., "Theory of Solid Propellant Ignition and Response to Pressure Transients," Bulletin of the 19th Interagency Solid Propulsion Meeting, CPIA Publication 18, Vol. III, pp. 95–127, Silver Spring, Md., July 1963 (Confidential).

## C. Propulsion System Analysis for Thrusters of Differing Performance Characteristics,

D. J. Norton

### 1. Introduction

This article outlines a theoretical performance analysis for a propulsion system being considered for outer-planetary orbiters. These missions can be performed with the use of multiple thrusters of different types, such as a solid propellant and monopropellant hydrazine motors. The objective is to determine the sensitivity of mission capabilities to the performance parameter of individual thruster types. Figure 12 schematically depicts a possible configuration that employs four vernier monopropellant motors and a multiple-start solid propellant motor. The equations describing the payload of a space vehicle after a propulsive maneuver are written in terms of the specific impulse  $I_{sp}$  and the mass fraction  $\lambda$  of the several rocket motors. From this information, an analysis of the sensitivity of payload to  $I_{sp}$  and  $\lambda$  has been obtained. An example of a current design problem is described and the results are presented.

### 2. Analysis

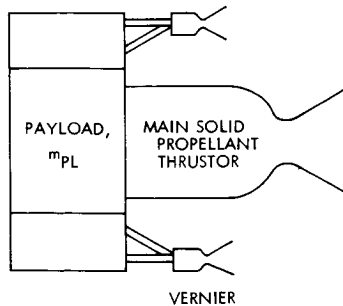
Consider a planetary probe mission that requires  $n$  separate  $\Delta V$  maneuvers during which no inert masses are staged. In general, the final mass of the spacecraft after the  $k$ th maneuver can be expressed in terms of the initial mass

$$M_{fk} = M_{0k} \exp(-\Delta V_k / I_{spk}) \quad (1)$$

where

$M_{fk}$  = spacecraft mass after  $k$ th maneuver

$M_{0k}$  = spacecraft mass before  $k$ th maneuver



$\Delta V_k$  = velocity change required for  $k$ th maneuver

$I_{spk}$  = delivered specific impulse

If no inert masses are ejected during the maneuver, the mass of propellant consumed is

$$M_{Pk} = M_{0k} (1 - B_k) \quad (2)$$

The spacecraft mass before the  $k$ th maneuver is,

$$M_{0k} = \prod_{j=0}^{k-1} B_j \quad (3)$$

where

$j = 0, \dots, k$  = index of maneuvers

$B_0 = M_{01}$  = initial spacecraft mass

$B_j = \exp(-\Delta V_j / I_{spj}) = M_{fj} / M_{0j}$

Combining Eqs. (2) and (3) yields

$$M_{Pk} = (1 - B_k) \prod_{j=0}^{k-1} B_j \quad (4)$$

The above results can now be extended to the entire sequence of  $n$  maneuvers by summation of  $k = 1, 2, \dots, n$ .

$$M_{fn} = \prod_{k=0}^n B_k \quad (5)$$

$$M_{PT} = \sum_{k=1}^n M_{Pk} = \sum_{k=1}^n (1 - B_k) \prod_{j=0}^{k-1} B_j \quad (6)$$

where

$M_{PT}$  = total propellant consumed

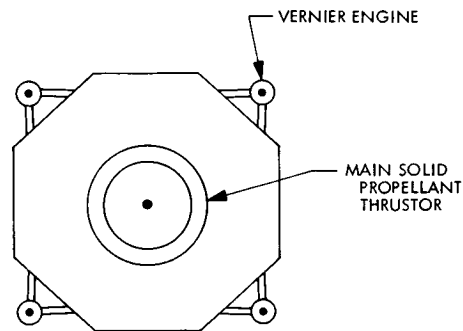


Fig. 12. Schematic of multiple thruster spacecraft

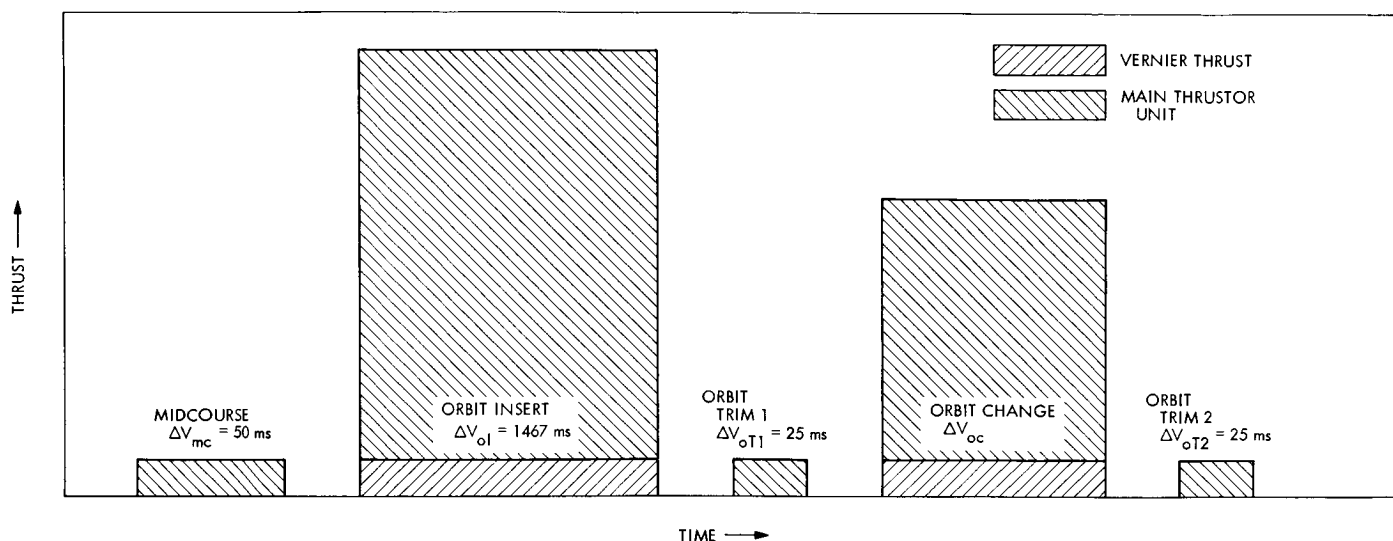


Fig. 13. Typical planetary orbiter mission profile

When more than one rocket type is employed during a maneuver, it is necessary to modify the form of the previous equation so that an average  $I_{sp}$  may be employed. The equations will be developed for two rocket types as depicted in Fig. 13; however, the analysis may be extended to account for more.

$$I_{sp} = \frac{(1 + \beta)I_{sp_a}}{\frac{I_{sp_a}}{I_{sp_b}} + \beta} \quad (7)$$

$$\beta = \frac{F_a}{F_b} = \left( \frac{\dot{m}_a}{\dot{m}_b} \right) \left( \frac{I_{sp_a}}{I_{sp_b}} \right) \quad (8)$$

where  $a$  and  $b$  are the vernier and main thruster, respectively, and  $\dot{m}$  is the propellant mass flow rate.

Equations (7) and (8) are strictly valid only when  $\dot{m}_a/\dot{m}_b = \alpha$  is a constant ratio during a particular maneuver. For the case when vernier units are fired together with a main unit, the ratio  $\alpha$  of mass flow rates of the motors should remain constant; however, their magnitudes may vary in fixed proportion.

In many analyses of potential propulsion systems, the prime variable of interest is the payload available after the last propulsive maneuver:

$$M_{PL} = M_{f_n} - (M_{Ia} + M_{Ib}) \quad (9)$$

where  $M_{Ia}$  and  $M_{Ib}$  are the inert masses of the vernier and main thruster, respectively.

It has already been shown how the final mass may be calculated; however, the inert masses of the rocket motors are more difficult to obtain.

$$M_{Ia} = \left( \frac{1 - \lambda_a}{\lambda_a} \right) M_{P_{Ta}} \quad (10)$$

$$M_{Ib} = \left( \frac{1 - \lambda_b}{\lambda_b} \right) M_{P_{Tb}} \quad (11)$$

where

$\lambda$  = mass fraction of the rocket motor

$M_I$  = inert rocket motor mass

$M_{P_T}$  = propellant consumed

the propellant used for the  $k$ th maneuver is

$$M_{Pa} = (1 - B_k) \left( \frac{\alpha_k}{1 + \alpha_k} \right) \prod_{j=0}^{k-1} B_j \quad (12)$$

$$M_{Pb} = \left( \frac{1 - B_k}{1 + \alpha_k} \right) \prod_{j=0}^{k-1} B_j \quad (13)$$

where

$\alpha = \infty$  when vernier only is firing

$\alpha = 0$  when main unit only firing

$\alpha = \frac{F_a I_{sp_b}}{F_b I_{sp_a}}$  when main and verniers are firing

The total mass of propellants consumed are

$$M_{PTa} = \sum_{k=1}^n \alpha_k \left( \frac{1 - B_k}{1 + \alpha_k} \right) \prod_{j=0}^{n-1} B_j \quad (14)$$

$$M_{PTb} = \sum_{k=1}^n \left( \frac{1 - B_k}{1 + \alpha_k} \right) \prod_{j=0}^{n-1} B_j \quad (15)$$

Therefore, the inert weights of the rockets may be computed as follows:

$$M_{Ia} = \left( \frac{1 - \lambda_a}{\lambda_a} \right) \sum_{k=1}^n \alpha_k \left( \frac{1 - B_k}{1 + \alpha_k} \right) \prod_{j=0}^{n-1} B_j \quad (16)$$

$$M_{Ib} = \left( \frac{1 - \lambda_b}{\lambda_b} \right) \sum_{k=1}^n \left( \frac{1 - B_k}{1 + \alpha_k} \right) \prod_{j=0}^{n-1} B_j \quad (17)$$

The useful payload may then be computed by applying Eqs. (5), (16), and (17) in Eq. (9).

$$M_{PL} = \prod_{j=0}^n B_j - (M_{Ia} + M_{Ib}) \quad (18)$$

### 3. Sensitivity Analysis

Often, after a preliminary design is determined, it is useful to ascertain how changes in the motor performance characteristics would improve the payload  $M_{PL}$ . This can be accomplished by finding the derivatives of  $M_{PL}$  with respect to  $I_{sp}$  and  $\lambda$  of the motors.

The derivatives for these changes may be analytically derived when only one type of rocket motor is considered.

$$\frac{M_{PL}}{M_0} = 1 - \left[ 1 - \exp \left( - \frac{\Sigma \Delta V}{I_{sp}} \right) \right] \cdot \frac{1}{\lambda} \quad (19)$$

$$\frac{\partial M_{PL}}{\partial I_{sp}} = \frac{M_0 \Sigma \Delta V}{I_{sp}^2 \lambda} \exp \left( - \frac{\Sigma \Delta V}{I_{sp}} \right) \quad (20)$$

$$\frac{\partial M_{PL}}{\partial \lambda} = M_0 \left[ 1 - \exp \left( - \frac{\Sigma \Delta V}{I_{sp}} \right) \right] \cdot \frac{1}{\lambda^2} \quad (21)$$

To determine if a fixed percent increase in  $I_{sp}$  will yield a greater payload than the same increase in  $\lambda$ , the following ratio should be evaluated:

$$\frac{I_{sp} \frac{\partial M_{PL}}{\partial I_{sp}}}{\lambda \frac{\partial M_{PL}}{\partial \lambda}} = \frac{x}{(e^x - 1)} \quad (22)$$

where

$$x = \frac{\Sigma \Delta V}{I_{sp}} \quad (23)$$

When more complicated missions are considered, it is convenient to perform numerical differentiation. This is accomplished by evaluating  $\Delta M_{PL}$  for a small change in  $I_{sp}$  or  $\lambda$ .

$$\frac{\partial M_{PL}}{\partial I_{sp}} = \frac{\Delta M_{PL}}{\Delta I_{sp}} \quad \frac{\partial M_{PL}}{\partial \lambda} = \frac{\Delta M_{PL}}{\Delta \lambda} \quad (24)$$

### 4. Results and Conclusions

To illustrate the usefulness of the analysis, an example problem was chosen that is of interest to the Jupiter orbiter study. Consider the mission that is illustrated schematically in Fig. 13. The spacecraft under consideration has four vernier motors plus a stop-restartable solid propellant motor for the large  $\Delta V$  maneuvers. The verniers alone are fired for the initial midcourse correction  $\Delta V_{mc}$ . This is followed by an orbit insert maneuver  $\Delta V_{oi}$ . For this maneuver both the verniers and the solid propellant motor are employed. Next, the verniers are used to provide an orbit trim  $\Delta V_{ot1}$ . Then a second solid propellant motor is fired in conjunction with the verniers to provide a change in the initial orbit  $\Delta V_{oc}$ . Finally, there is a second orbit trim  $\Delta V_{ot2}$  where only the vernier units contribute.

The maximum  $\Delta V_{oc}$  compatible with a payload  $M_{PL}$  of 1435 lbm is sought. The initial mass for the mission is limited by the launch vehicles from 2500 to 4500 lbm. Table 2 lists several candidate solid propellant options to be evaluated.

A computer program was written to evaluate propulsion candidates. The program computes the payload after

Table 2. Solid propellant options

Solid propellant motors	$I_{sp}$	$\lambda$
Conventional solid	290	0.92
Advanced solid	320	0.90
Hybrid solid	390	0.85
Vernier thrusters	$I_{sp}$	$\lambda$
Monopropellant hydrazine	250	0.80

the last propulsive maneuver. The partial derivatives for changes of payload with respect to  $I_{sp}$  and  $\lambda$  were also obtained.

For these studies,  $\beta = 0.05$  which is a representative value for the thrust ratio of the verniers compared to the solid motors to provide control of the spacecraft during main unit firing.

Figure 14 presents  $\Delta V_{oc}$ , which allows  $M_{PL} = 1435$  lbm as a function of  $M_0$  for the three solid options. The  $\Delta V_{oc}$  for these options are relatively close because of the trade-off of decreasing  $\lambda$  with increasing  $I_{sp}$ . Figure 15 presents the partials of  $M_{PL}$  with respect to  $I_{sp}$  and  $\lambda$ . A percentage increase of  $I_{sp}$  of the solid units is relatively more important than the same change in the vernier units. Also presented in Fig. 15 is  $\partial M_{PL}/\partial \lambda$  for the verniers and main thruster. Figure 16 presents the ratio of the partials for change of  $M_{PL}$  for the various options. A percentage change  $\lambda$  is relatively more important than  $I_{sp}$  for both the main thruster and the verniers units.

The usefulness of the results presented in Figs. 15 and 16 are best described by an example: Suppose the nominal solid propellant motor for a certain mission required that  $I_{sp} = 320$  lbf-s/lbm, and  $\lambda = 0.90$  to obtain the desired payload for an initial mass of 3100 lbm. It is desirable to use another propellant with  $I_{sp} = 305$  lbf-s/lbm because of the ease of manufacture. The question is then, what  $\lambda$  will be necessary for the new motor?

$$\Delta M_{PL} = \frac{\partial M_{PL}}{\partial I_{sp}} I_{sp} \left( \frac{\Delta I_{sp}}{I_{sp}} \right) + \frac{\partial M_{PL}}{\partial \lambda} \lambda \left( \frac{\Delta \lambda}{\lambda} \right) = 0 \quad (25)$$

$$\Delta \lambda = \frac{-\lambda \Delta I_{sp}}{I_{sp}} \left( I_{sp} \frac{\partial M_{PL}}{\partial I_{sp}} \right) \left( \frac{\lambda \partial M_{PL}}{\partial \lambda} \right)^{-1} \quad (26)$$

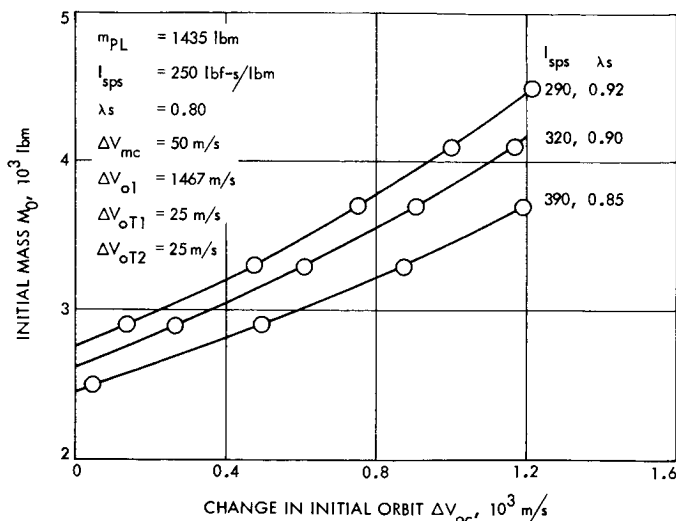


Fig. 14. Allowable  $\Delta V_{oc}$  as a function of payload for solid options

Examination of Fig. 16 reveals that for  $M_0 = 3100$  lbm the ratio of the partials for change is approximately 0.82.

$$\Delta \lambda = -(0.90) \frac{-15}{320} (0.82) = 0.0345 \quad (27)$$

Therefore, the new motor would have to have a mass fraction of at least  $\lambda = 0.9345$  to compete with the nominal motor.

The present analysis yields the expected performance of the unit for several propulsion candidates. In addition, it provides information on the most rewarding performance parameters for possible improvement; however, the theoretical values must, of course, be tempered by judgments as to the development costs and risks for the desired change.

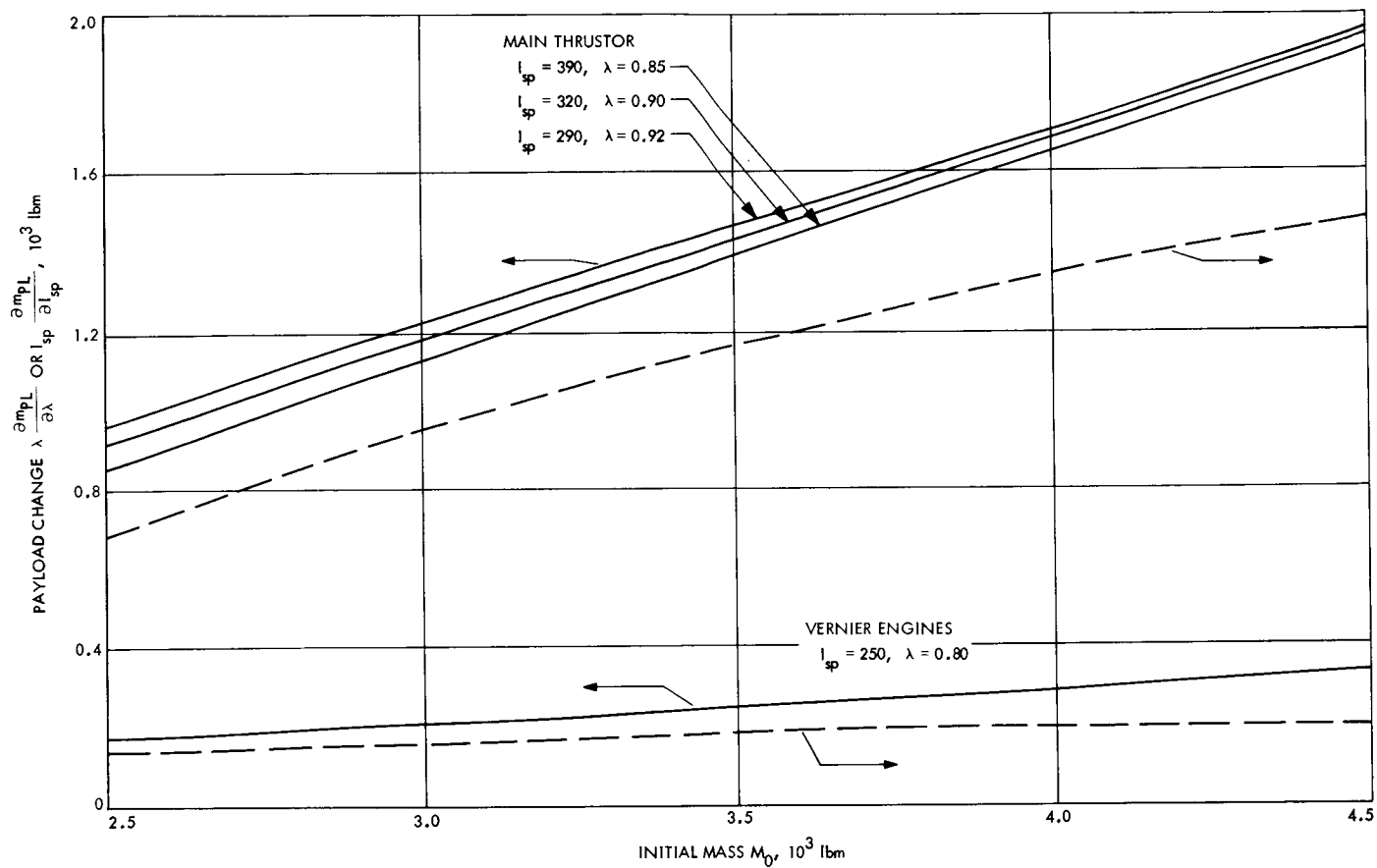


Fig. 15. Payload sensitivity to  $l_{sp}$  and  $\lambda$

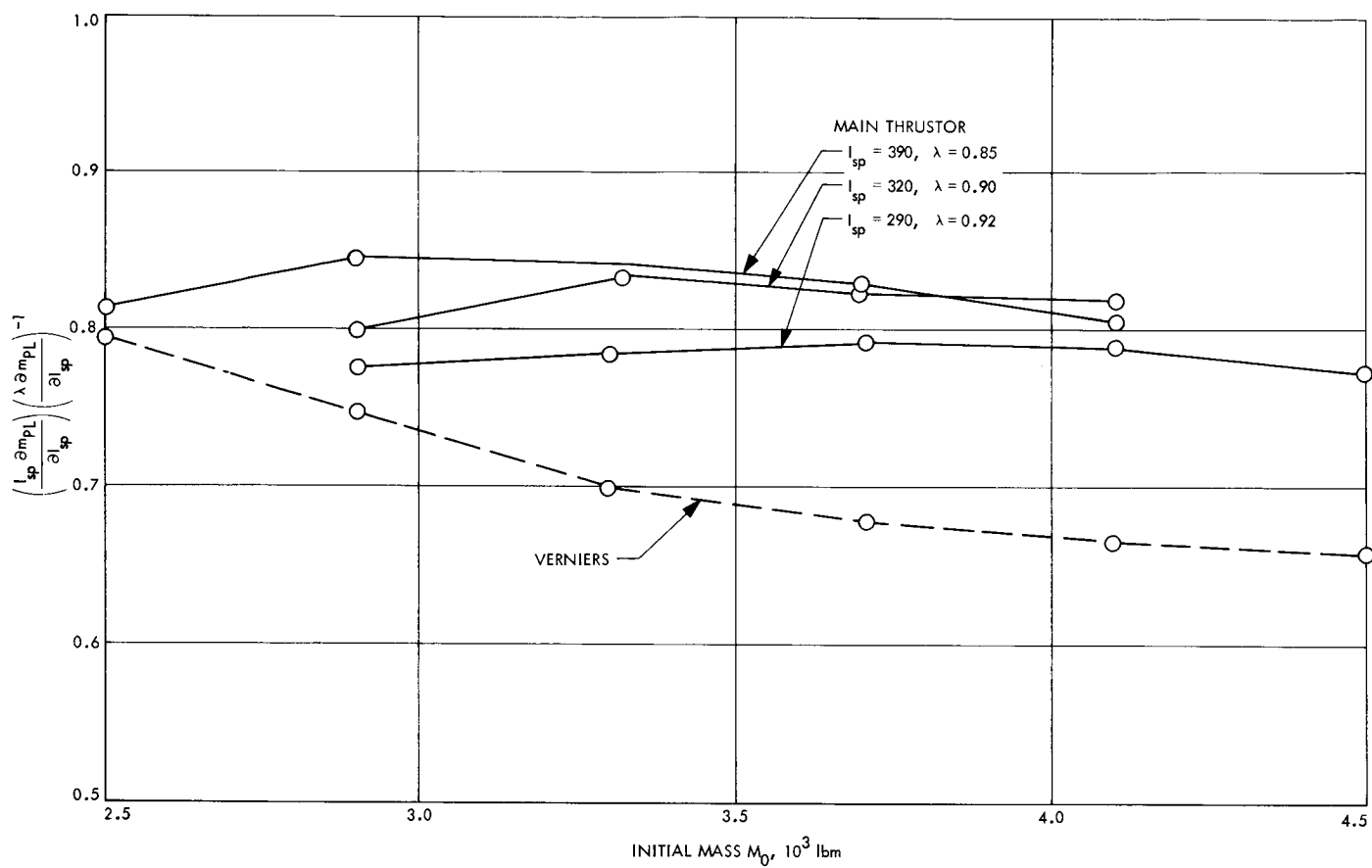


Fig. 16. Ratio of payload sensitivity for  $I_{sp}$  and  $\lambda$

## XIII. Polymer Research

### PROPULSION DIVISION

#### A. Investigation of Sterilizable Battery

**Separator Membranes,** E. F. Cuddihy,  
D. E. Walmsley, and J. Moacanin

##### 1. Introduction

Sterilizable battery separator membranes for silver-zinc batteries are currently prepared by grafting poly(potassium acrylate) onto thin films of polyethylene. The details of preparation and the chemical and physical analysis of the grafted membranes have been previously reported.<sup>1</sup>

These separator membranes are intended to prevent soluble silver oxide released by the silver electrode from contacting and contaminating the zinc electrode while permitting ready permeation of the KOH electrolyte solution. In practice, these membranes allow a low level of silver ion permeation and become black with time because of accumulation of a mixture of silver oxide and metallic silver (SPS 37-59, Vol. III). The membranes were also found to be losing a significant amount of grafted poly(potassium acrylate).

<sup>1</sup>SPS 37-50, Vol. III, pp. 166-169; SPS 37-52, Vol. III, pp. 98-100; SPS 37-54, Vol. III, pp. 100-104; SPS 37-56, Vol. III, pp. 186-189; SPS 37-59, Vol. III, pp. 200-204.

##### 2. Experimental Method

*a. Grafted membranes.* Two types of grafted membranes, prepared for JPL by the Southwest Research Institute, were employed in this investigation. One is the GX-152 series which is crosslinked with an aromatic crosslinker, divinylbenzene (DVB). The other is identical to GX-152 except that the crosslinker is 1,7-octadiene. Experimental verification of the presence of the material in the membrane is pending, but comparison can be made on the basis of the presence or absence of DVB in the membranes. Samples of these membranes were treated as follows:

- (1) Unsterilized samples of both membranes were exposed to a silver-oxide-saturated 46% KOH solution at 95°C.
- (2) Samples of GX-152 were first sterilized for 120 h in a 46% KOH solution at 135°C and then exposed to a silver-oxide-saturated 46% KOH solution at 95°C.
- (3) Unsterilized samples of GX-152 were exposed to a silver-oxide-saturated 46% KOH solution at 50°C.
- (4) Unsterilized samples of GX-152 were exposed to a 46% KOH solution at 95°C that contained no dissolved silver oxide.

The potassium acrylate content was determined for each sample before exposure; [because of inhomogeneous grafting (SPS 37-59, Vol. III), there was considerable concentration variability among specimens from the same membrane]. After exposure, the specimens were analyzed for final potassium acrylate content (SPS 37-54, Vol. III), and the total amount of deposited silver products along with the percentage of the deposit occurring as metallic silver and silver salts (SPS 37-59, Vol. III). The amount of deposited silver products were expressed in parts per hundred parts of the initial amount of potassium acrylate.

A procedure employed in the experimental scheme may result in erroneous estimates of deposited silver oxide. After exposure, the specimens are washed in water to remove any traces of the test solution. Calculations based on solubility constants and the size of our specimens indicate that the volume of wash water employed could dissolve up to 100 times the residual amounts of silver oxide. However, silver oxide deposited in the hydrophobic polyethylene phase should not be leached out. Therefore, the weight-to-weight ratios of silver oxide to polyethylene are given. The analytical procedure will have to be modified to determine the exact distribution at 95°C. Samples of two different grades of polyethylene phases.

**b. Polyethylene membranes.** To ascertain the relative roles of polyethylene and poly(potassium acrylate) in the deposition of silver products, samples of polyethylene were exposed to a silver-oxide-saturated 46% KOH solution at 95°C. Samples of two different grades of polyethylene were tested: (1) a totally saturated polyethylene film, Dow 400, and (2) Petrothane 301 which contained 0.12 mM/g of unsaturated pendent vinylidene groups.

### 3. Results and Discussion

**a. GX-152 membranes.** The results from exposure of the GX-152 membrane to 46% KOH solution at 95°C that contained no silver oxide are given in Table 1. After 1796 h, the membrane appears to rapidly approach an asymptotic value of about 20% of KOH-extractable potassium acrylate. Equivalent losses were observed for GX-152 samples sterilized 120 h in 46% KOH solutions, at 135°C, and additional exposure for 400 h at 95°C resulted in a further reduction of less than 3%. The observation of an apparent limiting solubility for both cases suggests that the KOH-extractable acrylate may be residual homopolymer left in the membrane rather than degraded grafted polymer.

**Table 1. Loss of potassium acrylate from unsterilized GX-152 membranes exposed to a 46% KOH solution at 95°C**

Value	Potassium acrylate, <sup>a</sup> %			
	21 h	162 h	812 h	1726 h
Initial	45	46	46	46
Final	44	42	41	40
Loss, % of initial	2	12	17	20
<sup>a</sup> Two membrane samples were used; one for the 21-h determination and a second for the other three. Exposure and monitoring of the second membrane was continued in excess of 1726 h, although these data are not currently available.				

The results for GX-152 membrane samples exposed to silver-oxide-saturated 46% KOH solution are given in Table 2. The sterilized materials (essentially free of KOH-extractable acrylate) experienced a 35% loss of acrylate after 800 h. The losses for the sterilized samples were 15% after 69 h and 22% after 316 h, which are lower than the losses observed for the unsterilized samples at equal times. The losses for the unsterilized samples are higher because they contain the initial KOH-extractable acrylate. With the information in Table 1 to account for the presence of the KOH-extractable acrylate, the losses caused by oxidative degradation are calculated to be 17% after 69 h and 22% after 316 h; values in reasonable agreement with the results for the sterilized sample.

The continuous loss of acrylate and simultaneous accumulation of metallic silver for both the sterilized and the unsterilized GX-152 membrane at 95°C are detailed in Table 2. Silver oxide dissolved in the KOH solution yields a strong oxidizing media that could attack the membrane and result in degradation and loss of the acrylate from the membrane. Simultaneously, the silver oxide would be reduced to metallic silver and deposited within the membrane. The equivalence ratio of potassium acrylate lost to metallic silver deposited is also given in Table 2. The equivalent weight ratios of acrylate to metallic silver are less than unity. This implies that no simple reaction scheme could be proposed. Of course, some oxidative attack of the polyethylene could have occurred and indeed small quantities of metallic silver were found in exposed polyethylene samples (Table 3).

Table 2 also contains a single result obtained after 528 h at 50°C for an unsterilized GX-152 membrane. The trends are identical to those obtained at 95°C, including the expected result that potassium acrylate is

**Table 2. Results of exposure of GX-152 membranes to a silver-oxide-saturated 46 % KOH solution at 95°C and 50°C**

Item	Sterilized <sup>b</sup> (95°C)			Unsterilized <sup>c</sup> (95°C)				Unsterilized (50°C)
				Observed		Corrected for KOH-extractable acrylate		
	69 h	316 h	800 h	69 h	316 h	69 h	316 h	528 h
Potassium acrylate <sup>a</sup> , %	41 (45)	24 (29)	20 (28)	35 (41)	40 (47)	31 (36)	36 (42)	38 (47)
Potassium acrylate, % of initial value	15	22	35	22	33	17	22	23
Total gain of deposited silver products, wt % of initial potassium acrylate	21	104	57	27	49	33	68	7
Ag <sup>0</sup> , %	66	95	83	68	92	68	92	60
Ag <sup>+</sup> , %	34	5	17	32	8	32	8	40
Equivalence ratio, potassium acrylate lost/Ag <sup>0</sup> deposited	1.0	0.21	0.61	1.2	0.74	0.85	0.39	5.5
Weight ratio, g Ag <sup>+</sup> deposited/g polyethylene	0.059	0.018	0.037	0.059	0.033	0.059	0.033	0.025

<sup>a</sup>Separate test specimens were used for each exposure period. The number in parenthesis gives the percent for the particular specimen prior to exposure; this was necessary because of the large specimen-to-specimen variations caused by inhomogeneous grafting of poly(potassium acrylate) onto polyethylene (SPS 37-59, Vol. III).

<sup>b</sup>Sterilization for 120 h in 46% KOH solution at 135°C should remove the KOH-extractable potassium acrylate. Therefore, these losses should be due only to degradation by silver oxide. These and the corrected results should then be equivalent.

<sup>c</sup>Table 1 shows that a portion of the potassium acrylate is extractable by action of the KOH solution alone. Hence, observed losses include both the KOH-extractable acrylate and that degraded by the silver oxide. To account for the KOH extractable with the information in Table 1, Case 2 is a recalculation of the data so that the parameters reflect only the potassium acrylate lost by silver oxide degradation.

lost at a slower rate, here 23% after 528 h compared to 22% after 69 h at 95°C. The only anomalous parameter is the equivalence ratio value of 5.5. However, this figure may be inflated by speculating that the acrylate removal by KOH-extraction is proceeding at a faster rate than the silver-oxide oxidation. No parallel KOH-extraction data at 50°C is yet available.

**b. 1,7-octadiene-treated membranes.** The results for a 1,7-octadiene-treated grafted membrane are given in Table 4 for exposure at 95°C. It was suspected that the crosslinker DVB could attract and increase the deposition of silver oxide by introducing aromatic groups into the membrane. However, when the weight ratios of the base polyethylene (Petrothane 301) and the weight ratios given in Table 2 were compared, the DVB-free membrane did not yield any results significantly different from the membranes that contained DVB.

No parallel data are available at present on the acrylate losses in this system from KOH exposure at 95°C. Therefore, the high acrylate loss of 32% after only 69 h is not completely explainable. This is particularly more

**Table 3. Results of exposure of polyethylene films for 1000 h to a silver-oxide-saturated 46 % KOH solution at 95°C**

Deposited silver products	Polyethylene type	
	Dow 400	Petrothane 301
Wt %	2	4
Ag <sup>0</sup> , %	82	73
Ag <sup>+</sup> , %	14	27
Weight ratio, g Ag <sup>+</sup> /g polyethylene	0.15	0.025

curious because there is a further loss of only 2% acrylate at 316 h. There is also a noticeable reduction in the amount of deposited silver products compared to the DVB-containing GX-152 membranes (see Table 2). However, it would be premature to identify this with the differences in chemical features of the crosslinkers. Again, the equivalence ratio is high, presumably due to the inclusion of KOH-extractable acrylate in the reported losses.

**Table 4. Results of exposure of a 1,7-octadiene-treated membrane to a silver-oxide-saturated 46% KOH solution at 95°C**

Item	Exposure time	
	69 h	316 h
Potassium acrylate <sup>a</sup> , %	25 (35)	24 (34)
Potassium acrylate, <sup>b</sup> % of initial value	32	34
Total gain of deposited silver products, % of initial potassium acrylate	15	20
Ag <sup>0</sup> , %	60	60
Ag <sup>+</sup> , %	40	40
Equivalence ratio, potassium acrylate lost/Ag <sup>0</sup> deposited	3.5	2.8
Weight ratio, g Ag <sup>+</sup> deposited/g polyethylene	0.033	0.041
<sup>a</sup> See footnote a, Table 2.		
<sup>b</sup> Includes KOH-extracted and silver-oxide-degraded losses.		

*c. Polyethylene membranes.* The saturated Dow 400 membrane accumulated fewer silver products than the unsaturated Petrothane 301 (Table 3). As unsaturation is an easily oxidizable state, it was expected that the Petrothane 301 would accumulate more deposits, particularly metallic silver.

The amounts of silver oxide detected in these films are comparable to the amounts found in the grafted membranes when the results are expressed on a polyethylene basis. This observation certainly supports the possibility that the silver oxide detected in the grafted membranes are trapped in the hydrophobic polyethylene phase. If any silver oxide is removed by the wash water, it is either stripped from the polyethylene phase in some uniform reduction, or only from the acrylate phase of the grafted membrane.

## **B. Parallel-Plate Viscometer, R. F. Fedors and R. F. Landel**

### **1. Introduction**

In solid propellant manufacture, a critical parameter in both the mixing and casting stages is the viscosity of the oxidizer-binder system. The parallel-plate viscometer affords a simple and convenient experimental technique for evaluating the effect of variables such as filler content, filler particle size, and size distribution on the viscosity of slurries (Ref. 1). In the original device used at JPL, the slurry was placed at the center between two

large parallel circular plates, a constant load was applied to the movable top plate (the bottom plate was stationary), and the plate separation measured as a function of time (Ref. 1). The volume of slurry acted on by the applied load remains constant; i.e., none of the sample is lost during the test. The plate separation was monitored manually with a calibrated dial gage and well-known equations were then used to calculate the viscosity from the plate-separation-vs-time data (Ref. 1).

Subsequently, two devices were built that, for operational convenience, used linear voltage differential transformers to automatically and continuously monitor the plate separation.<sup>2</sup> One of these employs a spring system that serves to counterbalance the mass of the upper plate to provide an equilibrium and constant plate separation at the beginning of a test. The arrangement of the springs is such that the applied load and spring system is effectively aligned in series; this has the disadvantage that the load experienced by the sample is not constant but varies with the plate separation because part of the load is taken up by the extension of the springs. The relationship between viscosity, plate separation, and time have been derived and reported (Ref. 2) for cases where the load acting on the sample remains constant.

### **2. Statement of Problem**

*a. Constant load case.* A theory has been developed (Ref. 2) for the parallel-plate viscometer that uses the equation of motion of an incompressible fluid and is based on the following assumptions:

- (1) The sample is incompressible.
- (2) The sample is Newtonian.
- (3) No body forces act on the sample.
- (4) The motion is very slow.
- (5) There is no relative motion between the plates and the material in contact with the plates.
- (6) The plate separation  $h$  is small in comparison to the sample radius  $R$  ( $R > 10h$ ).

These assumptions lead to the relation

$$\frac{dh}{dt} = -\frac{2Fh^3}{3\pi R^4\eta} \quad (1)$$

<sup>2</sup>The devices were built by B. G. Moser and T. Sauer.

where  $F$  is the applied load,  $R$  is the radius of the sample, and  $\eta$  is the viscosity (Newtonian).

There are two basic ways in which the device can function that lead to two distinct equations relating  $\eta$  to  $h$  and time. In the first method of operation, the volume  $V$  of the sample (considered cylindrical in shape) is kept constant; i.e., none of the sample is lost from the plates. Since for a cylinder,  $R^4 = (V/\pi h)^2$ , then Eq. (1) can be written as

$$\frac{dh}{dt} = -\frac{2\pi F h^5}{3\eta V^2} \quad (2)$$

which for the case of constant  $F$  can be integrated directly to yield

$$\frac{Ft}{\eta} = \frac{3V^2}{8\pi} \left( \frac{1}{h^4} - \frac{1}{h_0^4} \right) \quad (3)$$

where  $h_0$  is the plate separation at time  $t = 0$ . A plot of  $\log Ft$  vs

$$\log \frac{3V^2}{8\pi} \left( \frac{1}{h^4} - \frac{1}{h_0^4} \right)$$

should yield a linear response with unit slope and from an intercept,  $\log \eta$  may be estimated. Extremely accurate measures of  $h$  are required especially for small values since the function in Eq. (3) contains  $h^4$ .

In the second mode of operation, the area  $a$  of the sample in contact with the plates is constant and equal to the radius of the plate. For a constant load, Eq. (1) is integrated directly to yield

$$\frac{Ft}{\eta} = \frac{3\pi a^4}{4} \left( \frac{1}{h^2} - \frac{1}{h_0^2} \right) \quad (4)$$

In this case, a plot of  $\log Ft$  vs

$$\log \frac{3\pi a^4}{4} \left( \frac{1}{h^2} - \frac{1}{h_0^2} \right)$$

is predicted to yield a linear response. An advantage of this mode of operation is indicated by a comparison of Eqs. (3) and (4). Whereas the constant volume mode involves  $h^4$ , the constant area mode is characterized by  $h^2$  and therefore is less sensitive to small errors in the estimation of  $h$ .

In an attempt to relax the restriction (Ref. 2) that  $R > 10h$ , Gent (Ref. 3) derived the following equation for a parallel-plate viscometer:

$$F = -\frac{3\eta V}{h^2} \frac{dh}{dt} - \frac{3\eta V^2}{h^5} \frac{dh}{dt} \quad (5)$$

The first term on the right-hand side represents the effect of compression due to the approach of one plate toward the other. The second term is equivalent to Eq. (1). For the case of constant volume of sample and constant load, Eq. (5) is readily integrated to yield

$$\frac{Ft}{3\eta V} = \left( \frac{1}{h} - \frac{1}{h_0} \right) + \frac{V}{8\pi} \left( \frac{1}{h^4} - \frac{1}{h_0^4} \right) \quad (6)$$

This differs from Eq. (3) by the inclusion of the term containing the first power of the plate separation. Equation (6) reduces to that of Eq. (3) when  $R > h$ ; if, in fact,  $R = 10h$ , the first term on the right-hand side contributes only about 2% of the second term.

**b. Variable load case.** A schematic diagram of the JPL spring viscometer is shown in Fig. 1. Because of unavoidable design considerations, there is an effective spring system acting in extension and in series with the sample. For a given applied constant load  $F$ , the downward motion of the top plate produces both an extension of the springs and a reduction in the height of the sample. Hence, part of the total  $F$  is used to extend the springs and part is used to cause the sample to flow. If the load-extension response of the spring system alone were known, that portion of the total load acting on the sample could be determined.

The load-extension response for the spring system was characterized as follows: (1) reference marks were applied to both top and bottom plates and (2) the springs were adjusted to provide the maximum plate separation ( $h = 0.660$  cm). At this separation, the mass of the top plate is supported entirely by the spring system. Calibrated weights were then added to the top plate and the separation measured by a traveling microscope. Data were taken in both the loading (decreasing  $h$ ) and unloading (increasing  $h$ ) modes.

The data are shown in Fig. 2 where the circles represent the response to loading while the crosses represent the response to unloading. It is clear that the system is linear and that there is no appreciable hysteresis in the

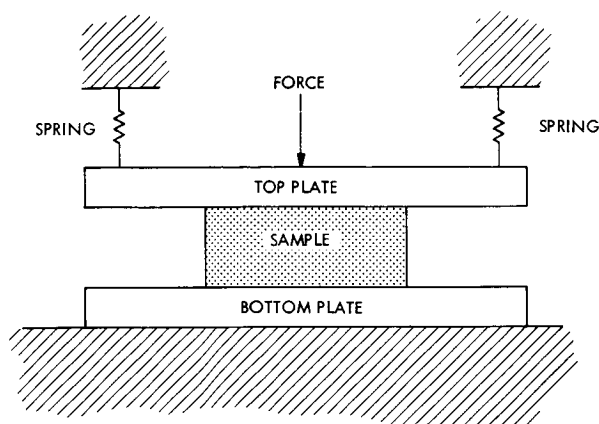


Fig. 1. Parallel-plate viscometer schematic diagram

spring system. Furthermore, the data in Fig. 2 represent the response of the system over the entire range of available plate separation and hence the system will also be linear over any other (smaller) interval of plate separation. The data may be represented by

$$F = -bh + c \quad (7)$$

where  $b$  is the slope ( $-195 \text{ g/cm}$ ) and  $c$  is the intercept at  $h = 0$  ( $c = 129 \text{ g}$ ). These numerical values for  $b$  and  $c$  are valid for an initial plate separation,  $h = 0.660 \text{ cm}$ . For any other value of  $h$ ,  $b$  will remain constant but  $c$  will decrease.

### 3. Discussion

If an experiment is carried out with a specimen between the plates when  $h = 0.660 \text{ cm}$ , and a constant load  $F_0$  is applied to the top plate, the effective load,  $F_{\text{eff}}$ , which acts to deform the sample is

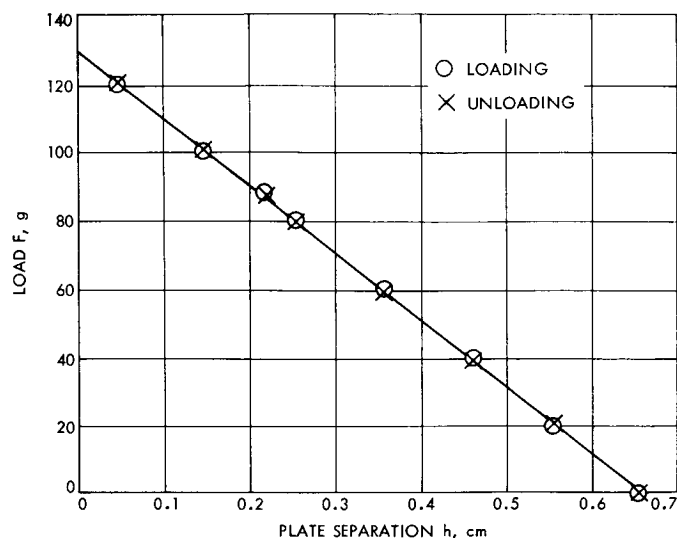


Fig. 2. Relationship between load and plate separation

$$F_{\text{eff}} = F_0 - F = F_0 + bh - c = f + bh \quad (8)$$

where  $f \equiv F_0 - c$ .

In Eq. (1),

$$\frac{dh}{dt} = -\frac{2Fh^3}{3\pi R^4 \eta} = -\frac{2h^3(f + bh)}{3\pi R^4 \eta} \quad (9)$$

For the constant volume case, Eq. (9) becomes

$$\frac{dh}{dt} = -\frac{2\pi h^5(f + bh)}{3\eta V^2} \quad (10)$$

which can be decomposed into partial fractions and integrated term-by-term to yield,

$$\frac{2\pi ft}{3V^2 \eta} \left[ \frac{1}{4} \left( \frac{1}{h^4} - \frac{1}{h_0^4} \right) - \frac{b}{3f} \left( \frac{1}{h^3} - \frac{1}{h_0^3} \right) + \frac{b^2}{2f^2} \left( \frac{1}{h^2} - \frac{1}{h_0^2} \right) - \frac{b^3}{f^3} \left( \frac{1}{h} - \frac{1}{h_0} \right) + \frac{b^4}{f^4} \ln \frac{\frac{f + bh}{h}}{\frac{f + bh_0}{h_0}} \right] \quad (11)$$

For the constant area case,

$$\frac{2ft}{3\pi a^4 \eta} = \frac{1}{2} \left( \frac{1}{h^2} - \frac{1}{h_0^2} \right) - \frac{b}{f} \left( \frac{1}{h} - \frac{1}{h_0} \right) + \frac{b^2}{f^2} \ln \frac{\frac{f + bh}{h}}{\frac{f + bh_0}{h_0}} \quad (12)$$

Similarly, by the use of Eq. (5), the following is obtained,

$$\frac{ft}{3\eta V} = \left\{ \frac{V}{2\pi} \left[ \frac{1}{4} \left( \frac{1}{h^4} - \frac{1}{h_0^4} \right) - \frac{b}{3f} \left( \frac{1}{h^3} - \frac{1}{h_0^3} \right) + \frac{b^2}{2f^2} \left( \frac{1}{h^2} - \frac{1}{h_0^2} \right) - \frac{b^3}{f^3} \left( \frac{1}{h} - \frac{1}{h_0} \right) + \frac{b^4}{f^4} \ln \frac{f+bh}{f+bh_0} \right] \right. \\ \left. + \left( \frac{1}{h} - \frac{1}{h_0} \right) - \frac{b}{f} \ln \frac{f+bh}{f+bh_0} \right\} \quad (13)$$

Although Eqs. (11–13) are more complex than their counterparts for constant load, all the parameters that appear are known or measurable and, therefore, can be used to estimate viscosity. It may be possible to simplify the equations somewhat, depending on the relative magnitude of  $f$ ,  $b$ , and  $h$ .

### References

1. Landel, R. F., Moser, B. G., and Bauman, A. J., "Rheology of Concentrated Suspensions: Effect of Surfactants," proceedings of the Fourth International Congress on Rheology, Part 2, held at Brown University, Providence, R. I. Aug. 26–30, 1963. Edited by E. H. Lee. Interscience Publishers, New York, 1965.
2. Dienes, G. J., and Klemm, H. F., *J. Appl. Phys.*, Vol. 17, p. 458, 1946.
3. Gent, A. N., *Brit. J. Appl. Phys.*, Vol. 11, p. 85, 1960.

## C. Relationship Between Index of Refraction and Molecular Weight, R. A. Rhein and D. D. Lawson

### 1. Introduction

This article demonstrates that the molecular weight of a polymer can be obtained from a relatively simple index of refraction measurement. Polymer molecular weight is ordinarily obtained from more laborious techniques such as viscosity, osmometry, light-scattering, or gel-permeation chromatography measurements.

An excellent linear correlation between index of refraction and inverse molecular weight for a series of poly(ethylene oxides) has been described by Ingham and Lawson (Ref. 1). This article shows that this correlation is as valid as the correlations between inverse molecular weight and other, more complex, functions of the index of refraction (both including and excluding density). This correlation is valid not only for the series

of poly(ethylene oxides), but also for other homologous series such as: saturated hydrocarbons, isobutylene telomers (up to heptaisobutylene), and nonconjugated dienes.

For each homologous series, a computer-assisted, least-squares linear fit was performed, correlating inverse molecular weight to the following parameters:

- (1) The index of refraction  $n$  (D line, 20°C)
- (2) A form of the Lorentz-Lorentz Equation (Ref. 2) is used where  $\rho$  is the density (i.e.  $D_4^{20}$ ).
  - (a)  $\frac{n^2 - 1}{n^2 + 2} \cdot \frac{1}{\rho}$
  - (b)  $\frac{n^2 - 1}{n^2 + 2}$
- (3) A form of the Eykman expression (Ref. 1),
  - (a)  $\frac{n^2 - 1}{n + 0.4} \cdot \frac{1}{\rho}$
  - (b)  $\frac{n^2 - 1}{n^2 + 0.4}$

It was found that for each series, the least-squares fit of  $n$  vs  $1/M$  (where  $M$  is the molecular weight) was as valid as the fit of  $1/M$  vs the other parameters, and generally superior to those including density.

### 2. Procedure

The molecular weight, refractive index, and density of each of the compounds in each of the four homologous series are shown in Table 5. For each of these groups, a least-squares linear fit of  $n$  vs  $1/M$ . The above-mentioned functions of  $n$  and density were performed on a computer, and in each case in a slope, intercept, and correlation coefficient were obtained.

Table 5. Molecular weight, refractive index, and density of four homologous groups

Homologous group	M	n	Density	Homologous group	M	n	Density
Poly (ethylene oxides) ( $n_D^{75}$ , $D_4^{75}$ ) <sup>a</sup>				Hepta-isobutylene	392.7	1.4739	0.8455
E 300	303.5	1.44592	1.080	Saturated hydrocarbons ( $n_D^{20}$ , $D_4^{20}$ )			
E 400	385.0	1.44776	1.078	2, 2-Dimethylbutane <sup>b</sup>	86.2	1.3675	0.6512
E 600	570.0	1.44958	1.080	2, 2, 4-Trimethylpentane <sup>b</sup>	114.2	1.39163	0.6918
E 1000	991.2	1.45121	1.082	2, 6-Dimethyldecane <sup>b</sup>	170.3	1.4228	0.7559
E 1450	1523	1.45202	1.077	5-Methyldodecane <sup>b</sup>	184.4	1.4244	0.7576
E 1540	1556	1.45227	1.080	4, 5-Dipropyloctane <sup>b</sup>	198.4	1.4322	0.7735
E 2000	20.8	1.45236	1.079	n-Hexadecane <sup>c</sup>	226.4	1.4352	0.7751
E 3000	3091	1.45311	1.079	n-C <sub>20</sub> H <sub>42</sub> <sup>c</sup>	282.6	1.4426	0.7887
E 4000	4545	1.45348	1.080	n-C <sub>24</sub> H <sub>50</sub> <sup>c</sup>	338.7	1.4480	0.7991
E 5000	5405	1.45354	1.080	n-C <sub>28</sub> H <sub>58</sub> <sup>c</sup>	394.8	1.4520	0.8067
E 6000	6350	1.45361	1.082	Non-conjugated dienes ( $n_D^{20}$ , $D_4^{20}$ ) <sup>b</sup>			
E 10000	9523	1.45374	1.082	2, 5-Hexadiene	82.1	1.4044	0.6883
Isobutylene telomers ( $n_D^{20}$ , $D_4^{20}$ ) <sup>b</sup>				2, 6-Octadiene	110.2	1.4292	0.7480
Di-isobutylene	112.2	1.4112	0.7195	2, 6-Dimethyl-1, 5-heptadiene	124.2	1.4400	0.7570
Tri-isobutylene	168.3	1.4310	0.7595	2, 6-Dimethyl-2, 6-nonadiene	152.3	1.4581	0.8007
Tetra-isobutylene	224.4	1.4482	0.7944	2, 6-Dimethyl-2, 6-decadiene	166.3	1.4594	0.8034
Penta-isobutylene	280.5	1.4588	0.8146	2, 5, 9-Trimethyl-4, 8-decadiene	180.3	1.4597	0.8037
Hexa-isobutylene	336.6	1.4684	0.8340				

<sup>a</sup>Reference 1.<sup>b</sup>Beilstein, *Handbook of Organic Chemistry*, Second Edition, Band 1, System 1-151. Springer-Verlag, Berlin, 1941.<sup>c</sup>Rossi, F. D., et al., "Selected Values of Physical and Thermodynamic Properties of Hydrocarbons and Related Compounds," API Research Project 44. Carnegie Press, Phila, Pa.

The correlation coefficient is a measure of the "goodness of fit" of the data. A correlation coefficient of zero indicates no correlation, or a totally random scattering of data; a correlation coefficient of either +1.000 or -1.000 indicates that all of the data fall on the line. The correlation coefficient is defined<sup>3</sup> as:

$$c = \frac{\sum_{i=1}^n (Y_i - \bar{Y}) \cdot (X_i - \bar{X})}{\left\{ \left[ \sum_{i=1}^n (Y_i - \bar{Y})^2 \right] \cdot \left[ \sum_{i=1}^n (X_i - \bar{X})^2 \right] \right\}^{1/2}}$$

where  $X_i$  and  $Y_i$  are the  $i$ th set of  $N$  observations, and  $\bar{X}$  and  $\bar{Y}$  are the means of the  $N$  values of  $X_i$  and  $Y_i$ .

A comparison of the correlation coefficients of  $n$  vs  $1/M$  and for the functions of  $n$  and  $\rho$ , for each group, indicates which function of  $n$  gives the "best fit" when plotted against  $1/M$ .

<sup>3</sup>"Concepts and Applications of Regression Analysis," IBM Corp., 1966.

The slopes, intercepts, and correlation coefficients are presented in Table 6. Upon examination, several features become apparent; with the exception of the isobutylene telomer group, the magnitude of the correlation coefficients for the functions that included density were somewhat less than the functions of  $n$  alone. Therefore, it is unnecessary to include density as a variable in the refractive-index molecular-weight correlation. Also, the magnitude of the correlation coefficient for the  $1/M$  vs  $(n^2 - 1)/(n + 0.4)$  and vs  $(n^2 - 1)/(n^2 + 2)$  were generally very slightly greater than that for the  $n$  vs  $1/M$  relationship for each homologous series, but the differences are essentially negligible. Hence,  $1/M$  correlates with  $n$  very nearly as well as with  $(n^2 - 1)/(n + 0.4)$  and  $(n^2 - 1)/(n^2 + 2)$ .

The slopes for vs  $n$   $1/M$  in Table 6 are nearly equal (with the exception of the poly(ethylene oxide) group) for the three other groups, and the values for the intercepts are a function of the degree of unsaturation. This is shown in Fig. 3, where the plots of  $n$  vs  $1/M$  for the

Table 6. Linear fit coefficients

Correlation Coefficients								
Function	Poly(ethylene oxides)		Isobutylene telomers		Saturated hydrocarbons		Non-conjugated dienes	
$n$	-0.99765		-0.98610		-0.99825		-0.99273	
$(n^2 - 1)/(n + 0.4)$	-0.99767		-0.98644		-0.99826		-0.99287	
$(n^2 - 1)/[(n + 0.4) \cdot d]$	-0.99427		0.99269		0.90528		0.91450	
$(n^2 - 1)/(n^2 + 2)$	-0.99754		-0.98713		-0.99815		0.99297	
$(n^2 - 1)/[(n^2 + 2) \cdot d]$	-0.95760		0.99130		0.95626		0.94536	
Slopes and intercepts								
Function	Poly(ethylene oxides)		Isobutylene telomers		Saturated hydrocarbons		Non-conjugated dienes	
	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept
$n$	-2.4094	1.4538	-9.8957	1.4954	-9.4539	1.4767	-8.9033	1.5122
$(n^2 - 1)/(n + 0.4)$	-3.009	0.6007	-12.348	0.6525	-11.879	0.6296	-11.129	0.6736
$(n^2 - 1)/[(n + 0.4) \cdot d]$	-2.5893	0.5561	3.2578	0.7327	1.7403	0.7364	3.9624	0.7329
$(n^2 - 1)/(n^2 + 2)$	-1.2455	0.2707	-5.1473	0.2923	-5.0278	0.2831	-4.6571	0.3012
$(n^2 - 1)/[(n^2 + 2) \cdot d]$	-1.0679	0.2506	2.026	0.3279	1.2863	0.3307	2.2915	0.3270

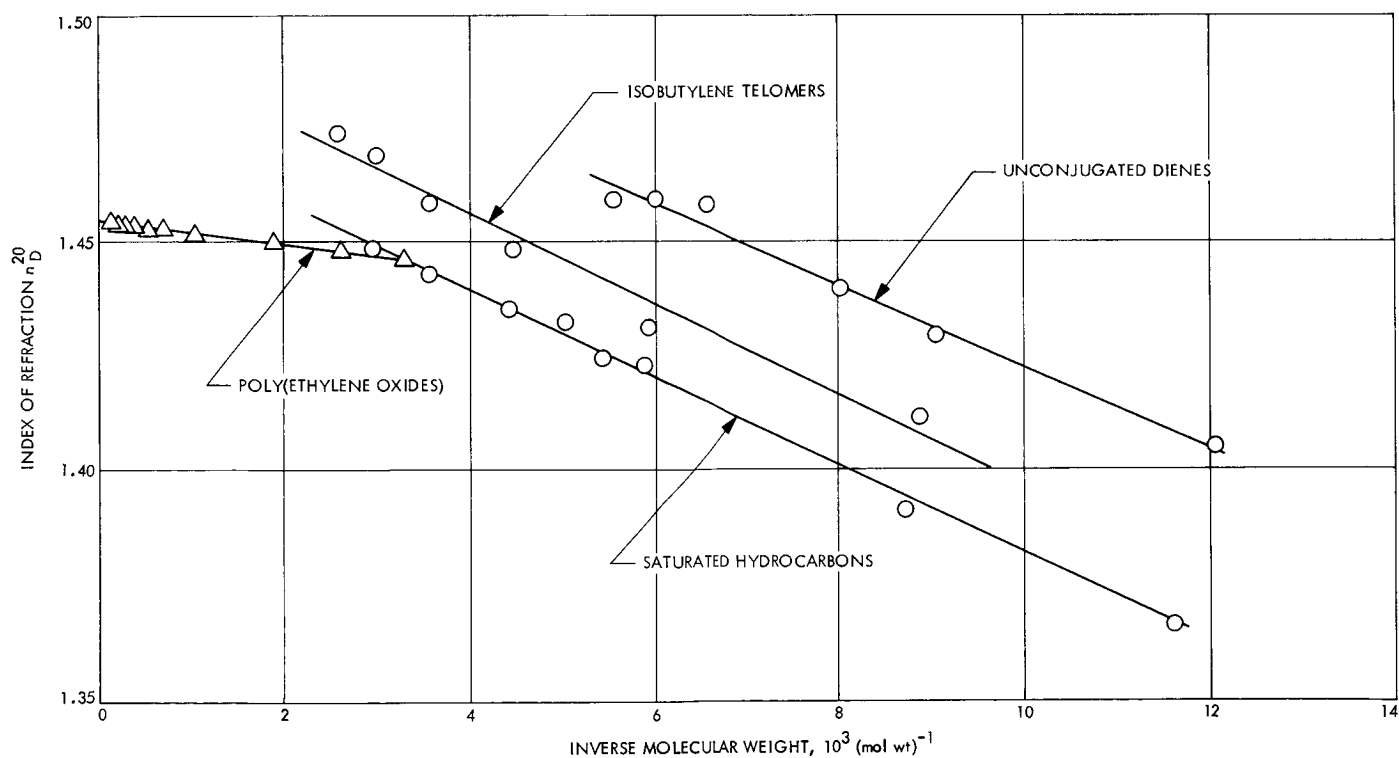


Fig. 3. Index of refraction vs inverse molecular weight

four groups, and the least-squares lines, are presented. Because the primary difference between the three groups is the degree of unsaturation, then measurements of  $n$  may be used to determine the degree of unsaturation.

### 3. Conclusions

For a homologous series of compounds, there is an excellent correlation between  $n$  and  $1/M$ , so that a good estimation of  $M$  can be made from an  $n$  measurement. Hence, density measurements are not necessary, nor is the use of expressions for  $n$  necessary.

For several homologous groups of compounds, each differing in the degree of unsaturation, the  $n$  vs  $1/M$  relationships show essentially parallel lines, with intercepts dependent upon the degree of unsaturation. Additionally, measurements of refractive index can also be used to determine the degree of unsaturation.

### References

1. Ingham, J. D., and Lawson, D. D., "Refractive Index-Molecular Weight Relationship for Poly(ethylene oxide)," *J. Polym. Sci.*, Vol. A3, No. 7, pp. 2707-2710, 1965.
2. Glasstone, S., *Textbook of Physical Chemistry*, p. 529. D. Van Nostrand Co., New York, 1959.

## D. Refractive Index to Molecular Weight Relationships for Isobutylene Telomers and Polymers, R. A. Rhein

### 1. Introduction

This article will show that good estimates of the number-average molecular weights (to  $\sim 1500$ ) of isobutylene telomers, telomer mixtures, and polymers can be obtained by relatively simple index of refraction measurements. Generally, more laborious methods, including changes in solvent colligative properties, viscosity relationships, or gel permeation chromatography, are used for molecular weight determination.

The results described here are from a least-squares linear fit of the index of refraction ( $n_D^{20}$ ) vs inverse molecular weight of a wide variety of isobutylene polymers and telomer mixtures. An excellent fit had been found previously (Ref. 1) for poly(ethylene oxide). In addition, it has been shown in *Chapter XIII-C* that density determinations are unnecessary in the estimation of molecular weight by the use of index of refraction measurements, and that the index of refraction correlates quite well

with inverse molecular weight for several homologous series of compounds.

### 2. Discussion

In this discussion, the indices of refraction ( $D$  line,  $20^\circ\text{C}$ ) were obtained either from literature or from direct measurement, for a wide variety of telomers, telomer mixtures, and polymers. The telomers were those in the homologous series di-isobutylene through hepta-isobutylene; the telomer mixtures were mixtures of di-, tri-, tetra-, and penta-isobutylene. Some of the polymers were commercially available polybutenes and others were prepared by the author by the molecular sieve catalyzed polymerization of isobutylene. These polymers often had wide molecular weight distributions, sometimes having two peak molecular weights (by gel permeation chromatography). These polymers were included to show the wide utility of this method for molecular weight determination.

Tables 7 and 8 show the isobutylene species considered here. The species are numbered (with few exceptions) in order of increasing molecular weight, and the index of refraction of each species is also listed. In Fig. 4, the index of refraction is plotted vs the inverse molecular weight for isobutylene polymers, telomers, and telomer mixtures. The line obtained by a computer-assisted, least-squares linear fit of refractive index vs inverse molecular weight is also shown in Fig. 4. The equation for the line corresponds to

$$n_D^{20} = 1.5034 - \frac{11.391}{M_n} \quad (1)$$

the correlation coefficient is  $-0.99254$ . From the equation for the line, the molecular weight can be computed from the equation

$$M = \frac{11.391}{1.5034 - n_D^{20}} \quad (2)$$

It was from Eq. (2) that the molecular weights were computed as shown in Tables 7 and 8. Also shown are the percent discrepancy (computed vs observed molecular weights) for each species. The average discrepancy, defined here as

$$\frac{100}{43} \cdot \sum_{i=1}^{43} \frac{|M_{c,i} - M_i|}{M_i} = 13.62\% \quad (3)$$

Table 7. Isobutylene telomers and telomer mixtures

Sample	Isobutylene, mol %				Observed index of refraction $n_D^{20}$	Number- average $M_n$ mol wt	Computed $M_c$ mol wt	Discrepancy, $100 \frac{M_c - M}{M}$ %
	Di-	Tri-	Tetra-	Penta-				
1	100.0	0.0	0.0	0.0	1.4112	112.2	123.5	+10.07
2	90.6	5.8	3.7	0.0	1.4116	119.6	124.1	+3.76
3	90.0	6.9	3.2	0.0	1.4124	119.6	125.1	+4.60
4	58.3	35.8	5.9	0.0	1.4216	138.9	139.2	+0.22
5	24.9	72.8	2.3	0.0	1.4293	155.6	153.7	-1.22
6	12.0	85.9	2.1	0.0	1.4310	162.7	157.3	-3.32
7	10.7	85.7	3.6	0.0	1.4305	164.3	156.2	-4.93
8	10.2	85.7	4.2	0.0	1.4320	165.0	159.5	-3.33
9 <sup>a</sup>	0.0	100.0	0.0	0.0	1.4310	168.3	157.3	-6.54
10	17.8	66.8	15.4	0.0	1.4348	167.0	166.0	-0.60
11	15.5	62.0	22.5	0.0	1.4390	172.3	176.8	+2.61
12	0.0	47.1	52.9	0.0	1.4441	198.0	192.0	-3.03
13	0.0	49.5	40.0	10.5	1.4474	199.4	203.3	+1.96
14	0.0	43.3	45.8	10.9	1.4466	203.0	200.5	-1.23
15	0.0	41.6	48.6	9.8	1.4464	203.7	199.8	-1.91
16	0.0	31.0	69.0	0.0	1.4464	207.0	199.8	-3.48
17 <sup>a</sup>	0.0	0.0	100.0	0.0	1.4482	224.4	206.3	-8.07
18 <sup>a</sup>	0.0	0.0	0.0	100.0	1.4588	280.5	255.3	-8.98

<sup>a</sup>Taken from Beilstein, *Handbook of Organic Chemistry*, Second Edition, Band 1, System 1-151. Springer-Verlag, Berlin, 1941.

Table 8. Isobutylene telomers and polymers

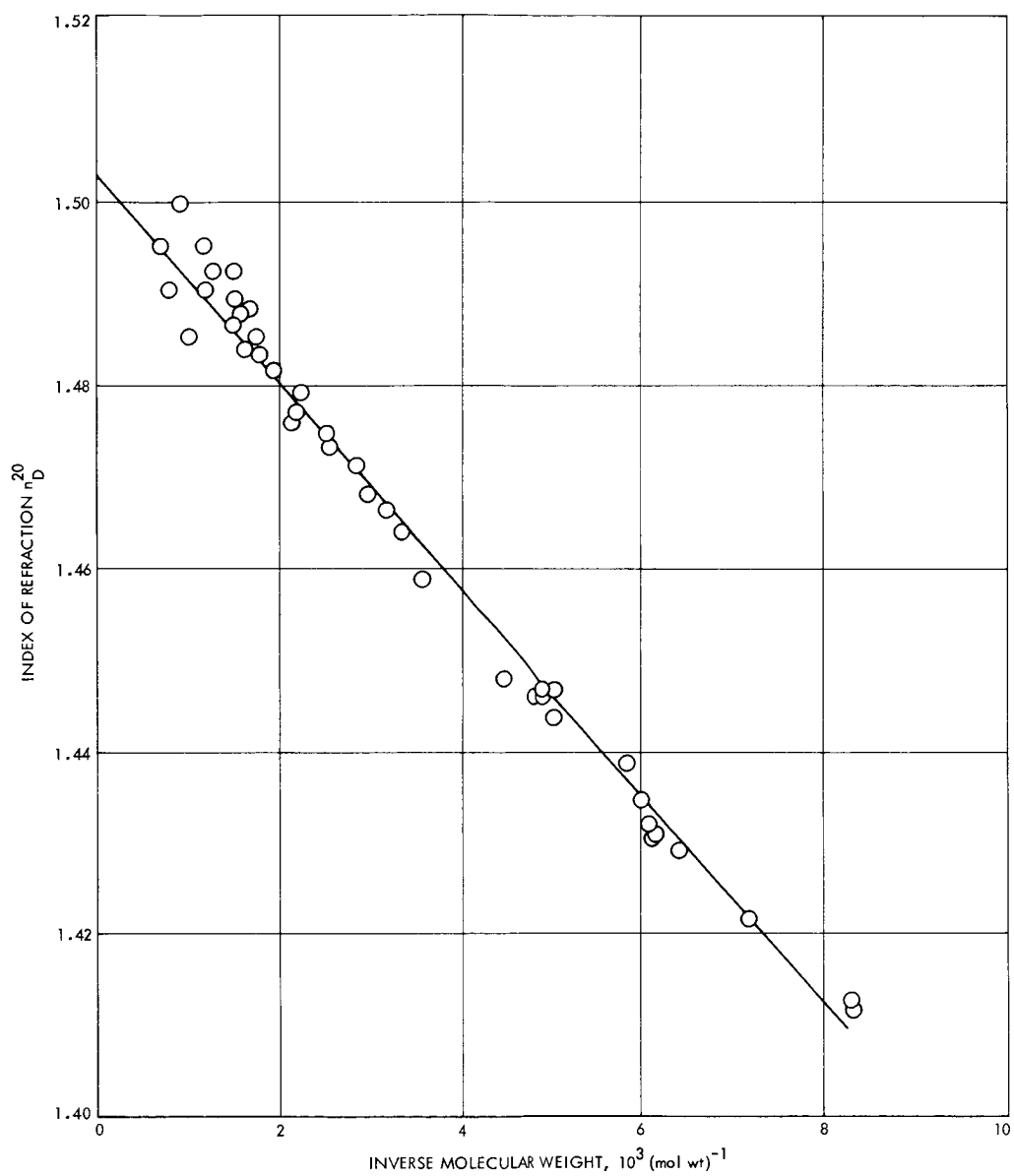
Sample	Description	Number-average $M$ , mol wt	Weight average, mol wt	Peak mol wt	Index of refraction $n_D^{20}$	Computed $M_c$ , mol wt	Discrepancy, $100 \frac{M_c - M}{M}$
19	Hexa-isobutylene <sup>a</sup>	336.6	336.6	336.6	1.4684	325.2	-3.39
20	Hepta-isobutylene <sup>a</sup>	392.8	392.8	392.8	1.4739	385.8	-1.78
21	10099-3rd	300	480	290	1.4641	290	-3.33
22	SK-100R	315	365	291	1.4669	312	-0.95
23	Polybutene 6 <sup>b</sup>	351	412	319	1.4714	356	+1.92
24	Polybutene 8 <sup>b</sup>	397	515	385	1.4750	401	+1.01
25	SK-400 R	448	2114	240, 3027	1.4794	474	+5.80
26	SK-310 R	457	761	351	1.4772	434	-5.03
27	9179-4	469	1397	277, 1302	1.4762	418	-10.87
28	9179-3	515	1264	291, 1302	1.4820	532	+3.30
29	9179-7	555	1325	305, 1186	1.4835	572	+3.06
30	PIB-685 <sup>c</sup>	572	715	645	1.4857	643	+12.41
31	SK-500 R	610	1608	404	1.4842	593	-2.79
32	Polybutene 16 <sup>b</sup>	635	1095	367	1.4875	715	+12.59
33	Polybutene 18 <sup>b</sup>	657	1150	510	1.4894	812	+23.59
34	9179-6	663	1963	302, 1725	1.4869	689	+3.92
35	Polybutene 20 <sup>b</sup>	693	1157	676	1.4924	1033	+49.06
36	SK-200 R	781	1232	560	1.4925	1043	+33.55
37	SK-45 R	839	2566	385, 3023	1.4908	902	+7.51
38	APS-1123-48.2 <sup>d</sup>	846	1533	580, 2020	1.4883	753	-10.99
39	Polybutene 24 <sup>b</sup>	847	1465	1030	1.4952	1385	+63.52
40	SK-110R	981	2164	535	1.4858	629	-35.83
41	Polybutene 32 <sup>b</sup>	1091	1976	1571	1.4998	3142	+187.9
42	11189-Z	1229	3438	487, 2888	1.4904	875	-28.80
43	PIB-1950 <sup>c</sup>	1416	2017	1894	1.4951	1368	-3.39

<sup>a</sup>Taken from Beilstein, *Handbook of Organic Chemistry*, Second Edition, Band 1, System 1-151. Springer-Verlag, Berlin, 1941.

<sup>b</sup>Courtesy of Mr. F. L. Delert, Chevron Chemical Co.

<sup>c</sup>Courtesy of Mr. R. Roper, ENSAY Polymer Laboratories.

<sup>d</sup>Courtesy of Dr. R. J. Lee, AMOCO Chemicals Corp.



**Fig. 4. Index of refraction vs inverse molecular weight for isobutylene polymers, telomers, and telomer mixtures**

### 3. Conclusion

In conclusion, good estimates of the molecular weight of isobutylene polymers, telomers, and telomer mixtures, for molecular weight <1500, may be obtained with the use of Eq. (2).

### Reference

1. Ingham, J. D., and Lawson, D. D., "Refractive Index-Molecular Weight Relationship for Poly(ethylene oxide)," *J. Polym. Sci.*, Vol. A3, No. 7, pp. 2707-2710, 1965.

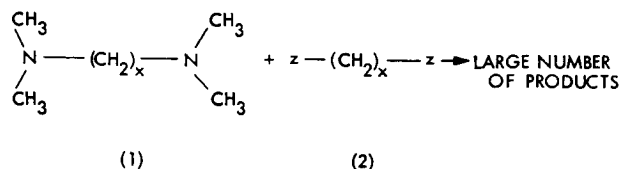
## E. Electronic Conductivity of Elastomeric Ionenes,

R. Somoano, S. P. S. Yen, and A. Rembaum

### 1. Introduction

The reaction between tetramethyl aminoalkanes and dihaloalkanes [(1) and (2), respectively, in Reaction 1] where  $x$  and  $y$  are integers and  $z$  is a halogen, leads to a wide variety of products according to the values of  $x$  and  $y$  (Ref. 1). With  $x < 3$  and  $y < 3$ , low molecular weight compounds are formed. With the correct choice of  $x$  and  $y$ , ionene polymers of weight-average molecular weight in the range of 10-40,000 may be synthesized.

However, all of the ionene polymers examined so far have consisted of crystalline materials exhibiting distinct



X-ray diffraction patterns. Their room temperature electrical resistivity, after reaction with LiTCNQ (TCNQ is tetracyanoquinodimethan) was found to be of the order of  $10^4$ - $10^7 \Omega\text{-cm}$  and their electronic transport properties will be reported separately.

Because of the great interest in conducting elastomers, Reaction (1) was applied to commercial "prepolymers" containing oxypropylene as a unit segment. The synthesis was based on the reaction of polypropylene glycol with tolylene diisocyanate and subsequent chain extension by means of a dimethylamino alcohol and a dihalide. This general reaction scheme (Fig. 5) permitted the preparation of a series of elastomeric ionenes, containing a variable concentration of positive charges located on quaternary nitrogen atoms. The different structure and the relatively low concentration of positive charges in these materials as compared with those synthesized previously (Ref. 1) are responsible for the rubbery properties. In this

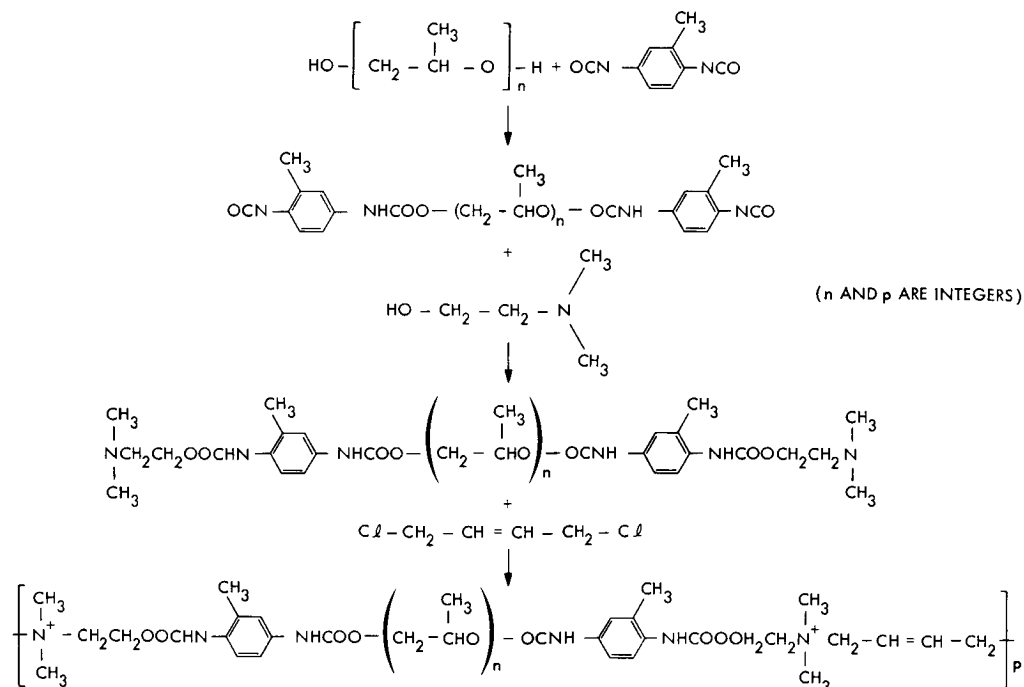


Fig. 5. Reaction sequence for the preparation of elastomeric ionenes

article, the electrical properties of elastomeric ionenes complexed with LiTCNQ are described. The study of electrical resistivity as a function of temperature revealed definite breaks in the resistivity temperature curves which could be related to the glass transition temperature of some of the materials.

## 2. Experimental

**a. Synthesis of elastomeric ionenes.** Dry dihydroxypolypropylene glycol of approximate molecular weight 2000 (100 g) was reacted with tolylene diisocyanate (17.5 g) at 60°C for 18 h. The reaction was followed by means of infrared spectroscopy. The disappearance of the OH peak at  $\lambda = 3480 \text{ cm}^{-1}$  ( $2.88 \text{ }\mu\text{m}$ ) indicated completion of the reaction.

After cooling the mixture to room temperature, 2-dimethylaminoethanol (DMAE) (8.9 g) was added while stirring. The reaction mixture was heated to 60°C for 2 h. The disappearance of the NCO absorption peak at  $\lambda = 2280 \text{ cm}^{-1}$  ( $4.38 \text{ }\mu\text{m}$ ) showed the end of reaction. After removal of excess DMAE in a vacuum evaporator, the dimethylamino content of the polymer was determined by an acid-base back titration method (Ref. 2).

The molecular weight of the dimethylamino end-capped prepolymer based on the amine titration was found to be 2514. Stoichiometric amounts of prepolymer and *trans* dichloro-2-butene were introduced into a glass vessel, stirred, sealed, and heated at 50°C for 24 h.

The purification consisted of precipitation of the polymer with a mixture of benzene and *n*-hexane from a methanol solution. After washing with *n*-hexane, the polymer was dried under vacuum at 50°C for 48 h. A similar procedure was applied to polypropylene glycol (PPG) of several molecular weights (1000, 2000, and 4000). For PPG 150 and 450, dry benzene was used as solvent to prepare dimethylamino end-capped prepolymer.

Addition of 1 mole LiTCNQ per mole of  $\text{N}^+$  to the polymer dissolved in methanol yielded a TCNQ complex which, after washing with water and drying, yielded elastomeric solids in the case of PPG 1000, 2000, and 4000. The different TCNQ-ionene complexes will be identified in this discussion by the corresponding molecular weight of PPG in the polymer.

**b. Electrical measurements.** Resistivity measurements (dc) were made on the TCNQ-ionene complexes. Sam-

ples that contained low molecular weight PPG ( $\leq 1000$ ) were compressed, *in vacuo*, into cylindrical pellets of 0.5-in. diam with a pressure of  $10^5$  psi. Samples that contained high molecular weight PPG (2000 and 4000) were pressed into sheets from which cylindrical samples were cut. Gold was evaporated onto the samples to insure good electrical contact. Resistivities were determined from current measurements with the use of a Keithley 610BR electrometer and a 95-V battery. The sample, placed in a teflon jig, was enclosed in a shielded cell that contained an inert gas and served as a temperature bath. The temperature was varied at approximately 1.5°C/min.

## 3. Results

The results of the dc resistivity measurements are shown in Figs. 6–8. The resistivities of all samples obey the Arrhenius relation

$$\rho = \rho_0 \exp\left(\frac{E_a}{kT}\right) \quad (2)$$

where  $\rho$  is the resistivity,  $\rho_0$  is the resistivity at infinite temperatures,  $k$  is the Boltzmann constant,  $T$  is the absolute temperature, and  $E_a$  is the activation energy. Values of the room temperature resistivity  $\rho_{25^\circ\text{C}}$  and the activation energy (in various temperature ranges) are given in Table 9.

A considerable effort was made to determine the reproducibility of the electrical data. Measurements were made on samples of the same compound, but from different batches, as well as on different samples from

Table 9. Activation energies and resistivities

Sample PPG, mol wt	Resistivity $\rho_{25^\circ\text{C}}$ , $\Omega\text{-cm}$	Activation energy $E_a$ , eV	Resistivity $\rho_{-65^\circ\text{C}}$ , $\Omega\text{-cm}$
150	$2.7 \times 10^7$	0.41	$2.8 \times 10^{10}$
450	$1.4 \times 10^8$	0.39 ( $T < +43^\circ\text{C}$ ) 0.99 ( $T > +43^\circ\text{C}$ )	$1.5 \times 10^{11}$
1000	$1.1 \times 10^8$	0.44 ( $T < -9^\circ\text{C}$ ) 0.63 ( $+25^\circ\text{C} > T > -9^\circ\text{C}$ ) 0.44 ( $T > +25^\circ\text{C}$ )	$4.0 \times 10^{11}$
2000	$2.2 \times 10^8$	0.47 ( $T < -38^\circ\text{C}$ ) 1.24 ( $+7^\circ\text{C} > T > -38^\circ\text{C}$ ) 0.80 ( $T > +7^\circ\text{C}$ )	$7.0 \times 10^{14}$
4000	$1.8 \times 10^8$	0.36 ( $T < -52^\circ\text{C}$ ) 0.91 ( $-3^\circ\text{C} > T > -52^\circ\text{C}$ ) 0.56 ( $T > -3^\circ\text{C}$ )	$4.0 \times 10^{14}$

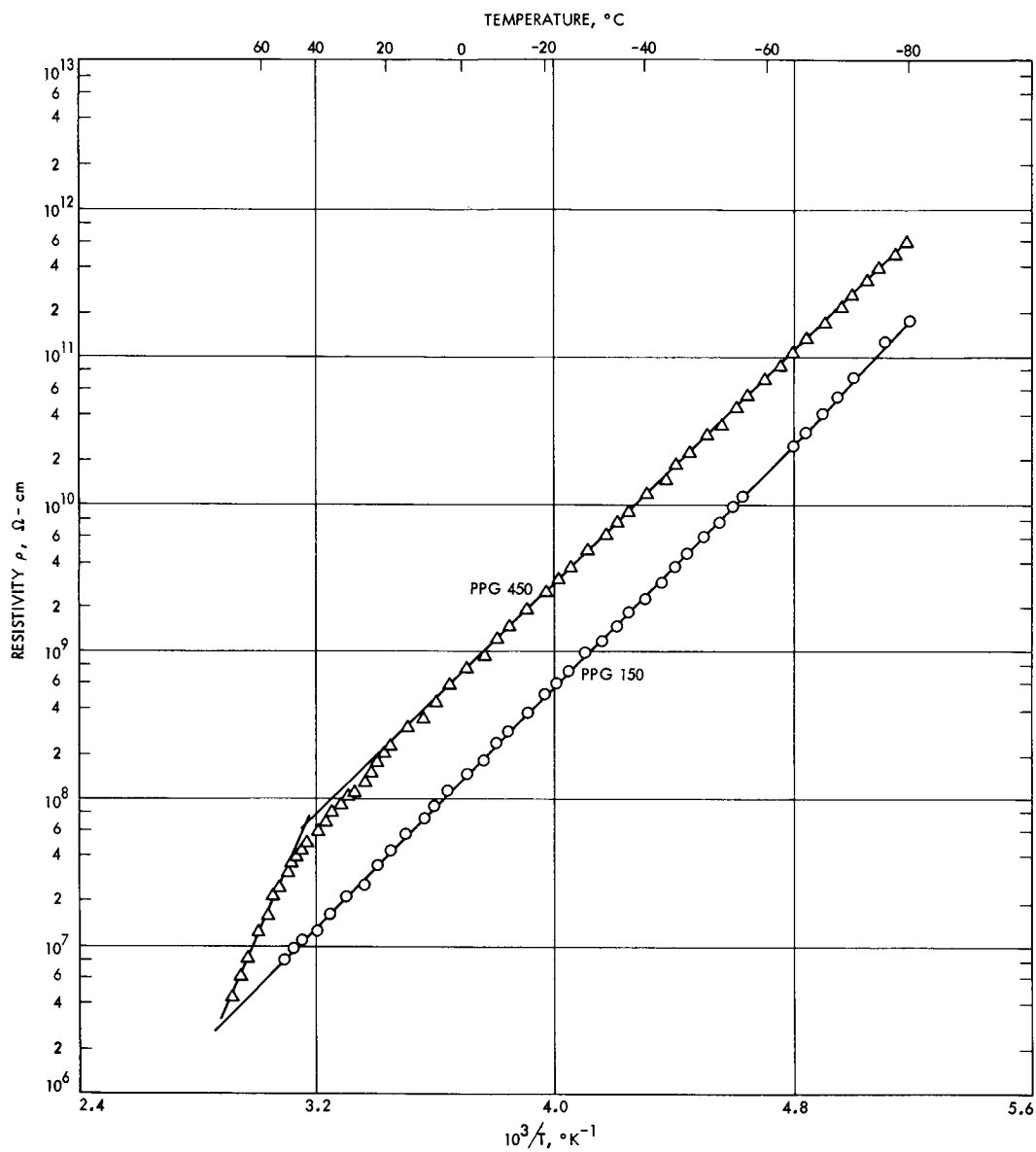


Fig. 6. Resistivity vs temperature for PPG 150 and 450

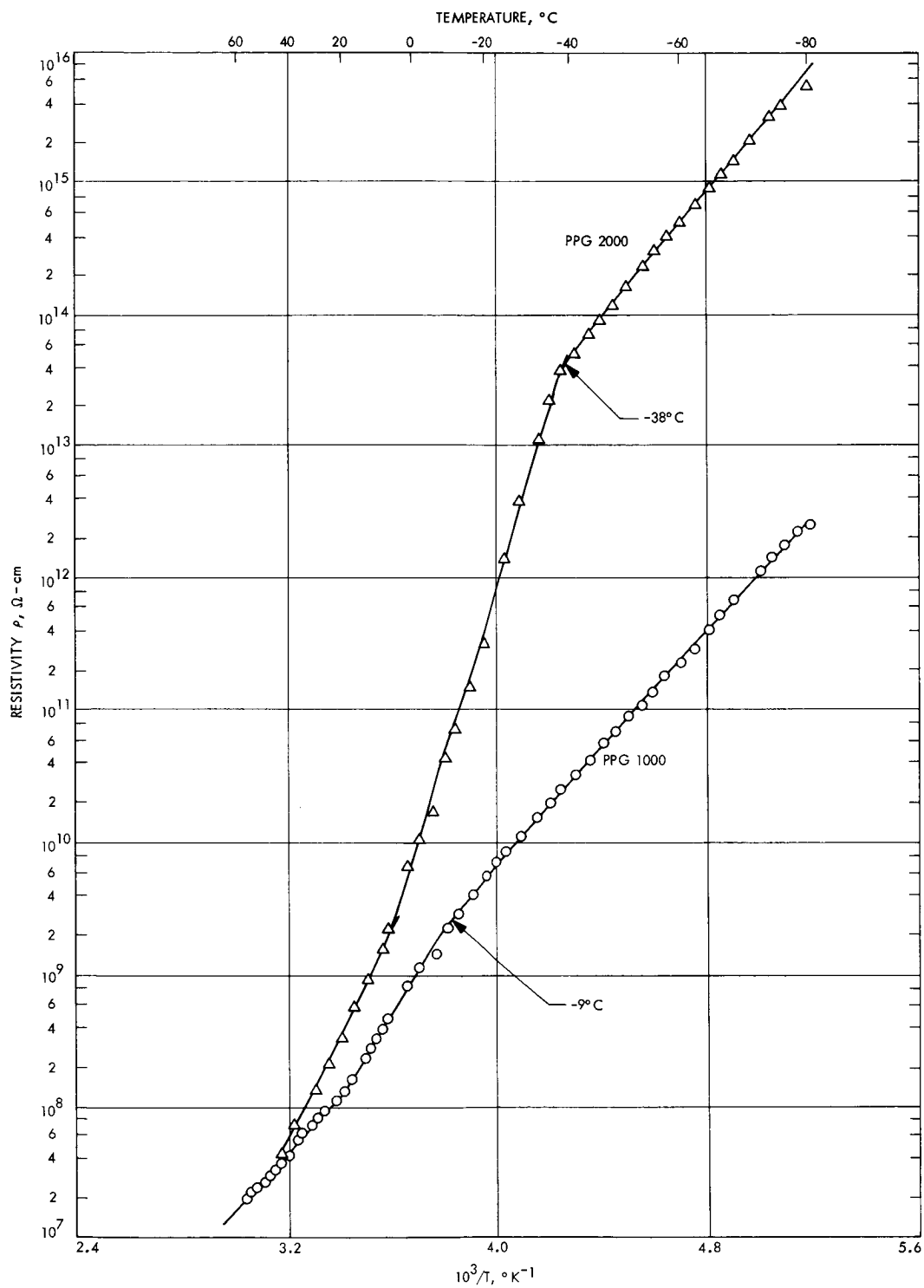


Fig. 7. Resistivity vs temperature for PPG 1000 and 2000

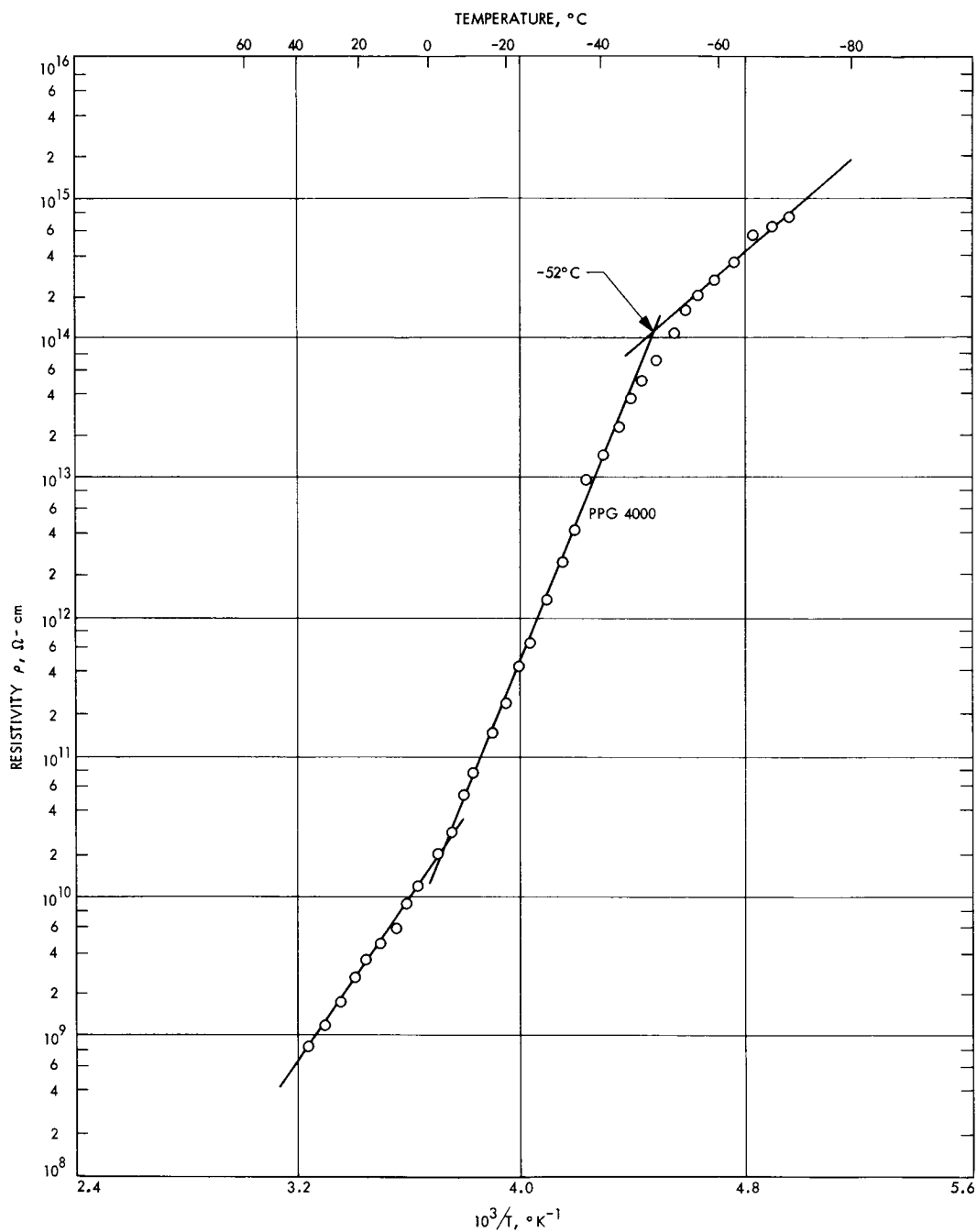


Fig. 8. Resistivity vs temperature for PPG 4000

the same batch. The reproducibility of the data for PPG 150, 450, and 2000 was excellent, as shown in Fig. 9. Measurements were made during both the heating and cooling cycles. However, on heating these samples to high temperatures ( $T > 80^{\circ}\text{C}$ ), the resistance, at constant temperature, increased rapidly with time, indicating that the samples were decomposing. The reproducibility of the PPG 4000 data was not as good. The physical properties of the material, i.e., rubbery particles, made it difficult to obtain good samples of suitable geometry and homogeneity. Figure 10 shows data taken on two samples of PPG 4000. The curves shown in Figs. 6-8 represent data taken on the best samples. The uncertainties in the  $\rho$  and  $E_a$  (Table 10) may be estimated from the reproducibility data of Figs. 9 and 10. Similar data for PPG 1000 were not taken, but the uncertainties in  $\rho$  and  $E_a$  should be between that of PPG 450 and PPG 2000.

The stability of the materials was checked by comparing UV-visible spectra of samples many weeks old with freshly prepared samples. The spectra were identical, indicating that the samples are stable over this time span. The room temperature resistivity values do not vary considerably as the molecular weight of PPG is increased, but the overall resistivity-temperature behavior changes significantly. To elucidate the structure exhibited in the resistivity-temperature curves of PPG 1000, PPG 2000, and PPG 4000, differential scanning calorimetry and differential thermal analysis tests were made on these samples. The results of these tests are shown in Figs. 11 and 12. A comparison of Figs. 6-8 with Figs. 11 and 12 indicates that the change in slope observed in the resistivity-temperature curves is associated with the glass transition in the low temperature region.

The glass transition has been observed in dc resistivity measurements by several workers (Refs. 3-5). However, in all of these cases, the current carriers were believed to be trace amounts of ionized impurities. It is believed

that the conduction processes in these TCNQ-ionene complexes are electronic (Refs. 6 and 7). The following observations substantiate this conclusion: (1) A current was passed through a sample of PPG 2000 for 17 days. The resistivity remained essentially constant, increasing less than a factor of 10 over this period. (2) For samples of low molecular weight ( $\text{PPG} \leq 1000$ ), the current response was instantaneous and constant upon application of the voltage, and reverse electromotive forces and currents were not observed on shorting the battery. In samples containing high-molecular-weight PPG, the current reached its steady-state value within 10 min, and the maximum reverse currents observed on shorting the battery were less than 1% of the previous steady-state value. (3) Saito (Ref. 5) has indicated that the pressure coefficient of resistance may be a useful criterion for establishing the nature of the current carriers. Figure 13 shows the room temperature resistance as a function of pressure of PPG 150 and PPG 450. It is observed that the pressure coefficient of resistance is negative. In systems believed to be electronic, such as inorganic glasses (Ref. 8), transition-metal oxide semiconductors (Ref. 9), and semiconducting organic polymers (Ref. 10), the pressure coefficient of resistance is also negative. The pressure coefficient of resistance is positive in ionic systems, such as silver chloride (Ref. 8) and various high resistivity polymers (Ref. 5).

#### 4. Conclusions

The resistivity is expected to be proportional to the distances between the TCNQ molecules associated with the positively charged nitrogen atoms on the ionene backbone. Changes brought about by varying the distances between TCNQ molecules may be determined by considering the weight-percent  $\alpha$  of the two charged nitrogen atoms in the repeatable unit. As  $\alpha$  increases or decreases, the distance between charges, as well as the resistivity, should decrease or increase, respectively. Figure 14 shows the variation of  $\alpha$  with the molecular weight of PPG. Corresponding effects in the resistivity can be seen clearly in the low-temperature region in which the various transitions have not yet occurred. In Table 9, values of the resistivity at  $-65^{\circ}\text{C}$ ,  $\rho_{-65^{\circ}\text{C}}$ , are given for comparison with Fig. 14. Upon initially increasing the molecular weight of PPG,  $\alpha$  decreases significantly and the resistivity similarly increases considerably. The value of  $\alpha$  varies only slightly on going from PPG 2000 to PPG 4000, and, similarly, the values of the resistivity of these two samples differ slightly. Therefore, the effect on the resistivity of varying the distances between charges is more pronounced in the low-molecular-weight samples ( $\text{PPG} \leq 1000$ ).

**Table 10. Uncertainties in resistivity  $\Delta\rho$  and activation energy  $\Delta E_a$**

Sample PPG, mol wt	Resistivity $\Delta\rho$	Energy $\Delta E_a$ , %
150	20%	5
450	20%	5
2000	Factor of 2	25
4000	Factor of 10	25

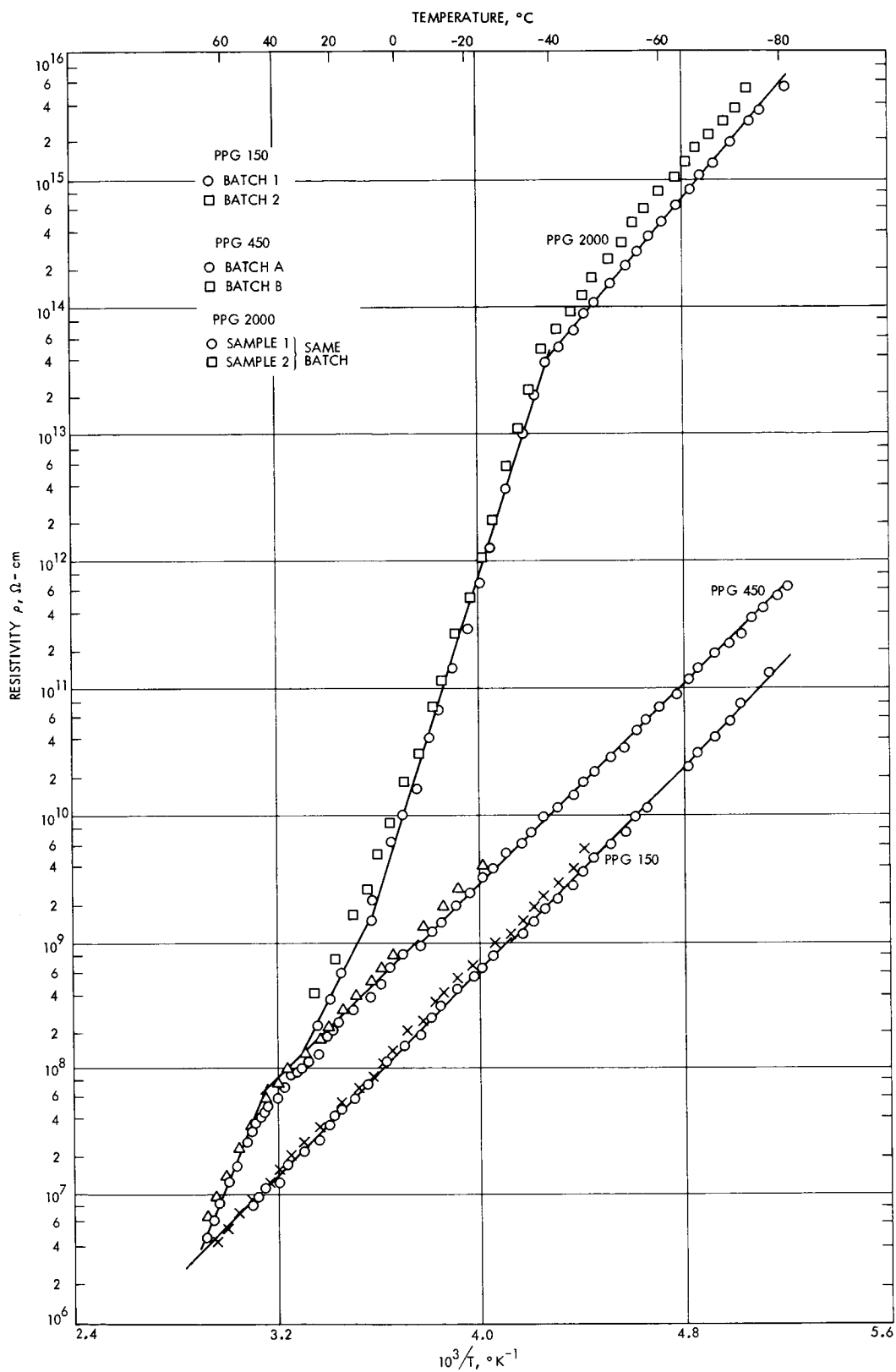


Fig. 9. Reproducibility of electrical data for PPG 150, 450, and 2000

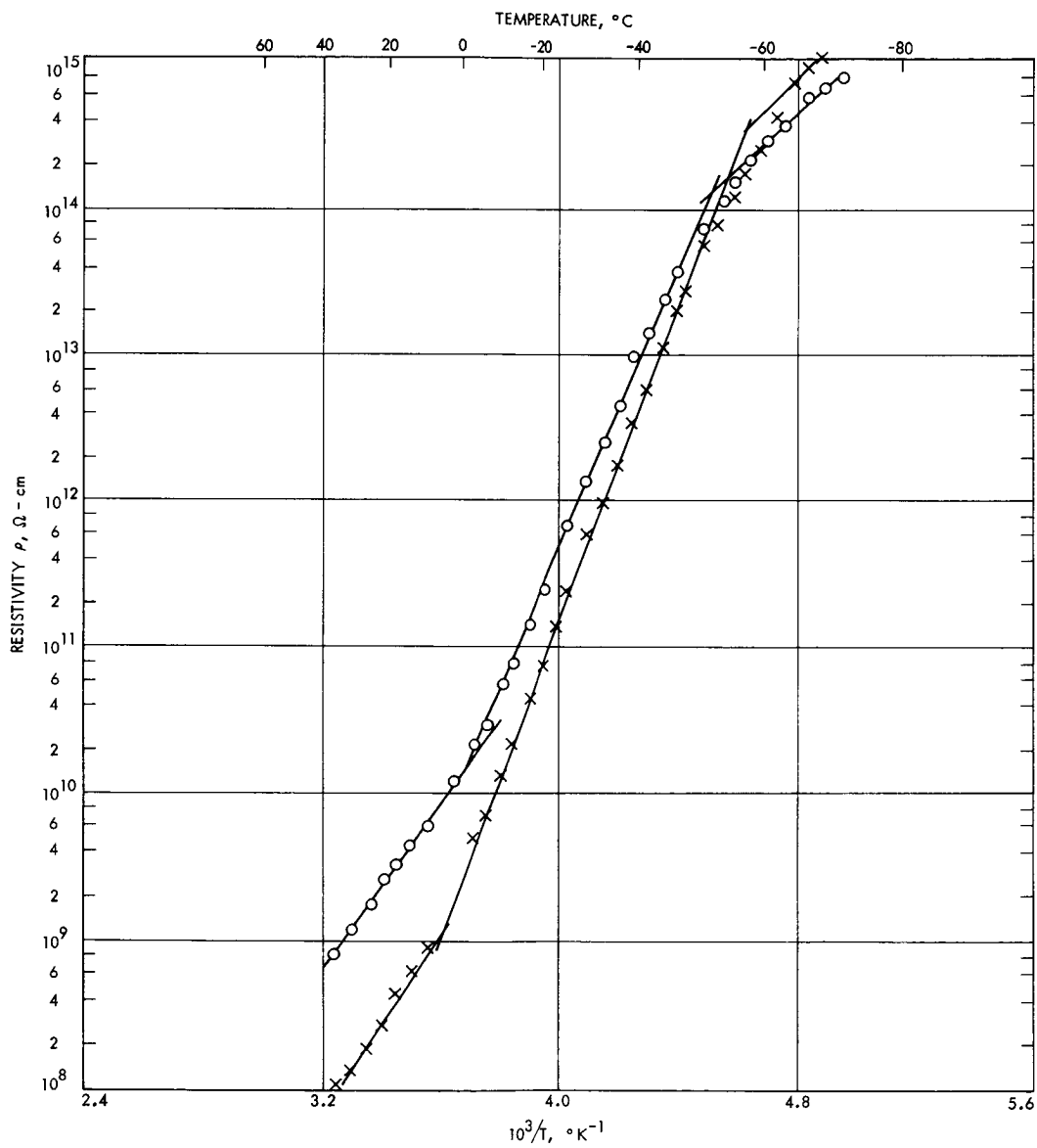
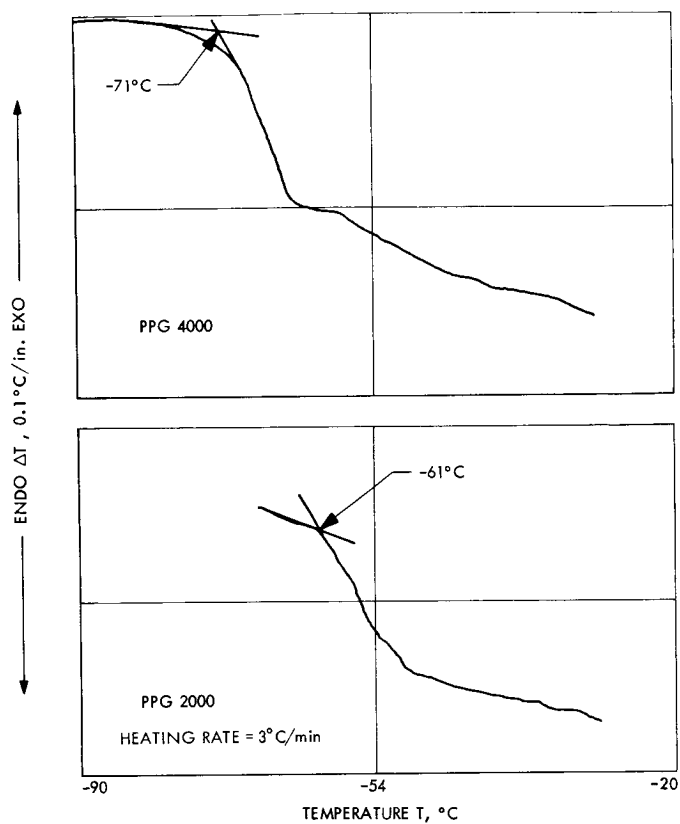
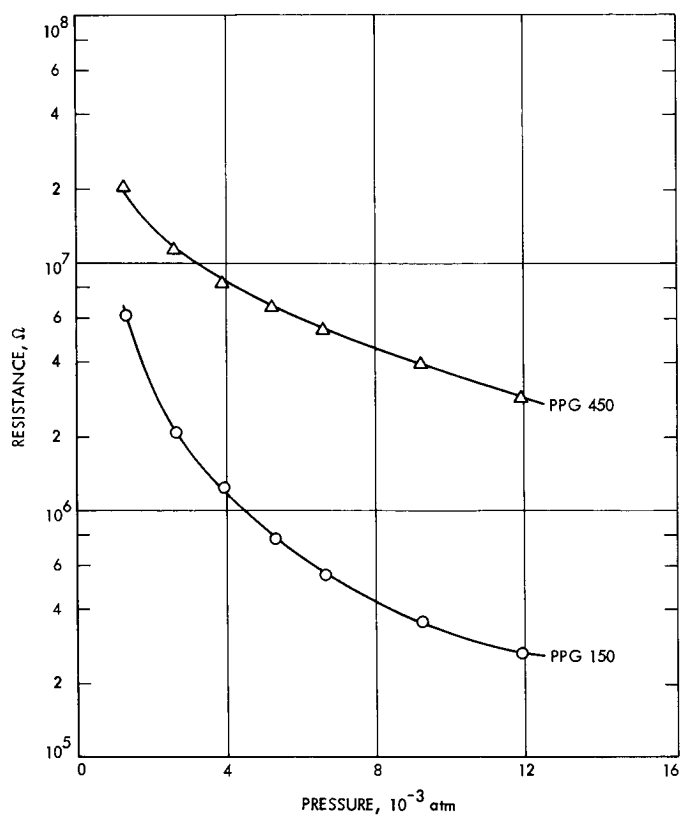


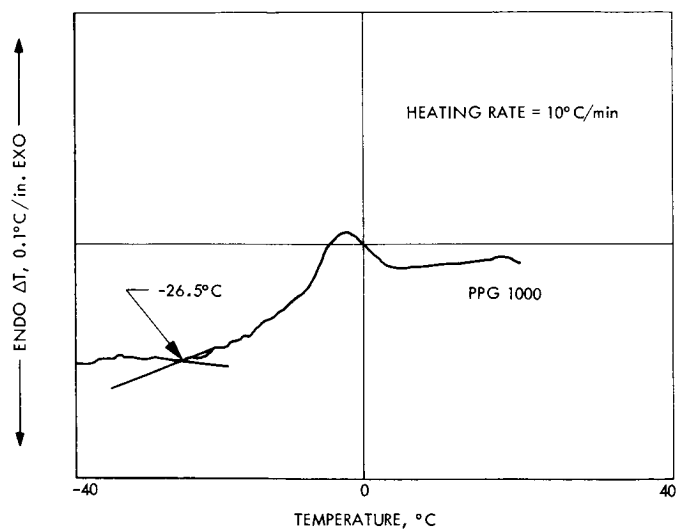
Fig. 10. Reproducibility of electrical data for PPG 4000



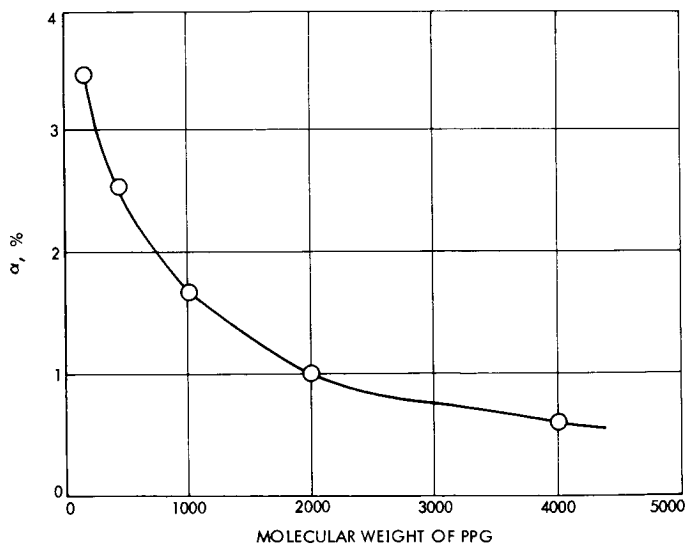
**Fig. 11. Differential thermal analysis test results for PPG 4000 and 2000**



**Fig. 13. Room temperature resistance vs pressure for PPG 150 and 450**



**Fig. 12. Differential scanning calorimetry test results for PPG 1000**



**Fig. 14. Percent by weight of two charged nitrogen atoms in the repeatable unit vs molecular weight of PPG**

The electrical measurements on the TCNQ-PPG ionene complexes indicate that the conductivity in these compounds is electronic in nature. This conclusion is supported by the negative pressure coefficient and polarization experiments as well as by previous findings (Ref. 6).

The low-temperature break in the dc temperature-resistivity curves represent the first observation of  $T_g$  via dc electrical methods in an electronic system. The  $T_g$  of complexed and uncomplexed high-molecular-weight elastomers (5% of TCNQ) does not differ to any significant extent. This is not true in the PPG 1000 case where the  $T_g$  of complexed material (10% of TCNQ) is 17°C higher than of the uncomplexed material. The higher  $T_g$  is not surprising because a high concentration of a bulky substituent on the polymer chain is expected to raise the  $T_g$ . The concentration of TCNQ in the high-molecular-weight polymer is not sufficient to have an appreciable effect on the  $T_g$ .

The low-temperature breaks in the curves of Figs. 6-8 occur at a temperature that is approximately 20°C higher than the  $T_g$  determined by differential scanning calorimetry or differential thermal analysis methods. This indicates that the electrical response to the onset of molecular motions is much slower than in the thermal methods. The high-temperature breaks in Figs. 6-8 may be caused by other transitions, the origin of which remains to be explored.

## References

1. Noguchi, H., and Rembaum, A., *J. Polym. Sci., Part B*, No. 7, p. 383, 1969.
2. Akagi, K., Oae, S., and Murakami, M., *J. Am. Chem. Soc.*, Vol. 79, p. 3118, 1957.
3. Warfield, R. W., and Petree, M. C., *Makromol. Chem.*, Vol. 58, p. 139, 1962.
4. Shishkin, N. I., and Vershinina, M. P., *Sov. Phys.-Solid State*, Vol. 1, No. 5, p. 724, 1958.
5. Saito, et al., *J. Polym. Sci.*, Vol. A-2, No. 6, p. 1297, 1968.
6. Kepler, R. G., *Phonons and Phonon Interactions*. Edited by T. A. Bak., W. A. Benjamin, Inc., New York, 1964.
7. LeBlanc, O. H., *J. Chem. Phys.*, Vol. 42, p. 4307, 1965.
8. Mackenzie, J. D., *Physics of Non-Crystalline Solids*. Edited by J. A. Prins. North Holland, Amsterdam, 1965.
9. Minomura, S., and Drickamer, H. G., *J. Appl. Phys.*, Vol. 34, p. 3403, 1963.
10. Pohl, H. A., Rembaum, A., and Henry, A., *J. Am. Chem. Soc.*, Vol. 84, p. 2699, 1962.

## F. Viscoelastic Behavior of Elastomers

Undergoing Scission Reactions, J. Moacanin,

J. J. Aklonis, and R. F. Landel

### 1. Introduction

A differential equation has been derived previously,<sup>4</sup> that allows the viscoelastic properties of elastomers experiencing simultaneous physical relaxation and chain scission to be predicted. This article examines techniques for the arithmetic solution of this differential equation and discusses the physical significance of this treatment.

### 2. Discussion

The stress-relaxation modulus  $G(t)$  for an elastomer undergoing scission can be represented by:

$$G(t) = G_e [\nu(t)] \left[ 1 + \left( \frac{t + \bar{m}}{\tau a_\nu} \right)^{-m} \right] \quad (1)$$

where  $G_e$  is the equilibrium modulus corresponding to the crosslink density  $\nu$  at time  $t$ ;  $\tau$  is a parameter appropriate for the initial  $\nu$ ,  $a_\nu$  is a scaling factor depending on  $\nu(t)$ ,  $\bar{m}$  is a function of the extent of degradation and  $m$  is a constant for a given elastomer type. To calculate  $\bar{m}$  for zero-order kinetics for chain scission, the following differential equation must be solved (SPS 37-58, Vol. III, p. 201, Eq. 17):

$$\frac{-d\bar{m}}{dt} = \frac{k}{m\tau^m} t^{1+m} \left( 1 + \frac{\bar{m}}{t} \right)^{1+m} (1 - kt)^{mb-1} \quad (2)$$

where  $k$  is the rate constant. For other reaction orders, the last term in the equation would be modified. Subject to several approximations that are reasonable for very low extents of reaction, Eq. (2) may be directly integrated to yield (SPS 37-58, Vol. III, p. 201, Eq. 20)

$$\begin{aligned} -\bar{m} &= (1 + m)^{-1} \left( \frac{m\tau^m}{k} \right)^{1/(1+m)} Z^{*(2+m)/(1+m)} \\ &\times \int_0^1 Y^{1/1+m} \exp [Z^*(Y - 1)] dY \end{aligned} \quad (3)$$

where

$$Z^* = \frac{k}{m\tau^m} t^{1+m}$$

<sup>4</sup>SPS 37-57, Vol. III, pp. 169-172; SPS 37-58, Vol. III, pp. 199-201; SPS 37-59, Vol. III, pp. 198-200.

The numerical solution of Eq. (3) was carried out and it was found that the magnitude of  $\bar{m}$  rapidly approached  $t$ . Because the variable of interest is  $t + \bar{m}$  ( $\bar{m}$  is a negative number),  $\bar{m}$  must be known accurately for the calculated value of  $t + \bar{m}$  to have any significance. The desired degree of accuracy was very tediously obtained and involved differences of very large numbers; therefore, a different evaluation scheme was formulated. Direct integration of Eq. (1) yields:

$$\bar{m}(t) = \frac{-k}{m\tau^m} \int_0^t (\xi + \bar{m})^{1+m} (1 - k\xi)^{mb-1} d\xi \quad (4)$$

The integration is then carried over a set of  $(j-1)$  time intervals defined by the sequence

$$t_1 = 0, t_2, t_{i-1}, t_i, t_{i+1}, \dots, t_{j-1}, t_j = t$$

Equation (3) then becomes:

$$+\bar{m}_j = \frac{-k}{m\tau^m} \sum_{i=1}^{i=j-1} \int_{t_i}^{t_{i+1}} (\xi + \bar{m})^{1+m} (1 - k\xi)^{mb-1} d\xi \quad (5)$$

where  $\bar{m}_i \equiv \bar{m}(t_i)$ . The summation can be written as:

$$m_j = -\frac{k}{m\tau^m} \int_{t_{j-1}}^{t_j} (\xi + \bar{m})^{1+m} (1 - k\xi)^{mb-1} d\xi - \frac{k}{m\tau^m} \sum_{i=1}^{i=j-2} \int_{t_i}^{t_{i+1}} (\xi + \bar{m})^{1+m} (1 - k\xi)^{mb-1} d\xi \quad (6)$$

According to Eq. (5), the sum is merely the definition of  $\bar{m}_{j-1}$ . A recursion relationship ensues:

$$\bar{m}_j = \bar{m}_{j-1} - \frac{k}{m\tau^m} \int_{t_{j-1}}^{t_j} (\xi + \bar{m})^{1+m} (1 - k\xi)^{mb-1} d\xi \quad (7)$$

If the interval  $t_{j-1}$  to  $t_j$  is small,  $\bar{m}$  may be approximated as a linear function of  $t$ ,

$$\bar{m}(t) = a_j t + c_j \quad t_{j-1} < t < t_j \quad (8)$$

Substitution of Eq. (8) into Eq. (7) yields

$$\bar{m}_j = \bar{m}_{j-1} - \frac{k}{m\tau^m} \int_{t_{j-1}}^{t_j} [(1 + a_j)\xi + c_j]^{1+m} (1 - k\xi)^{mb-1} d\xi \quad (9)$$

Restricting these considerations to low degrees of reaction ( $kt \ll 1$ ) allows a further approximation:

$$\bar{m}_j = \bar{m}_{j-1} - \frac{k}{m\tau^m} \left(1 - k \frac{(t_j + t_{j-1})}{2}\right)^{mb-1} \int_{t_{j-1}}^{t_j} [(1 + a_j)\xi + c_j]^{1+m} d\xi \quad (10)$$

which can now be integrated analytically to give

$$\bar{m}_j = \bar{m}_{j-1} - \frac{k}{m\tau^m} \left[1 - k \frac{(t_j + t_{j-1})}{2}\right]^{mb-1} \left[ \frac{[(1 + a_j)t_j + c_j]^{2+m} - [(1 + a_j)t_{j-1} + c_j]^{2+m}}{(2 + m)(a_j + 1)} \right] \quad (11)$$

where

$$c_j = \bar{m}_{j-1} - a_j t_{j-1}$$

and

$$a_j = \frac{\bar{m}_j - \bar{m}_{j-1}}{t_j - t_{j-1}}$$

The term  $a_j$  is approximated by  $a_{j-1}$ ;

$$a_j \approx a_{j-1} = \frac{\bar{m}_{j-1} - \bar{m}_{j-2}}{t_{j-1} - t_{j-2}}$$

For the cases examined, further iterations were found unnecessary.

Equation (11) is a recursion relationship for  $\bar{m}_j$  in terms of  $\bar{m}_{j-1}$ . The starting value  $\bar{m}(t_2)$  and  $\bar{m}(t_3)$  are conveniently calculated with the relationship

$$\bar{m}(t) = -\frac{k}{m(2+m)\tau^m} t^{2+m} \quad (12)$$

which holds for  $\bar{m} \ll t$  (i.e., very low extents of reaction, where  $-\bar{m}/t \ll 1$  and the  $(1 - kt)$  term can be neglected.

Because no data are known to exist with which to test the accuracy of results predicted by this technique, a parametric calculation was carried out. Parameters that characterize both relaxation and reactions were chosen

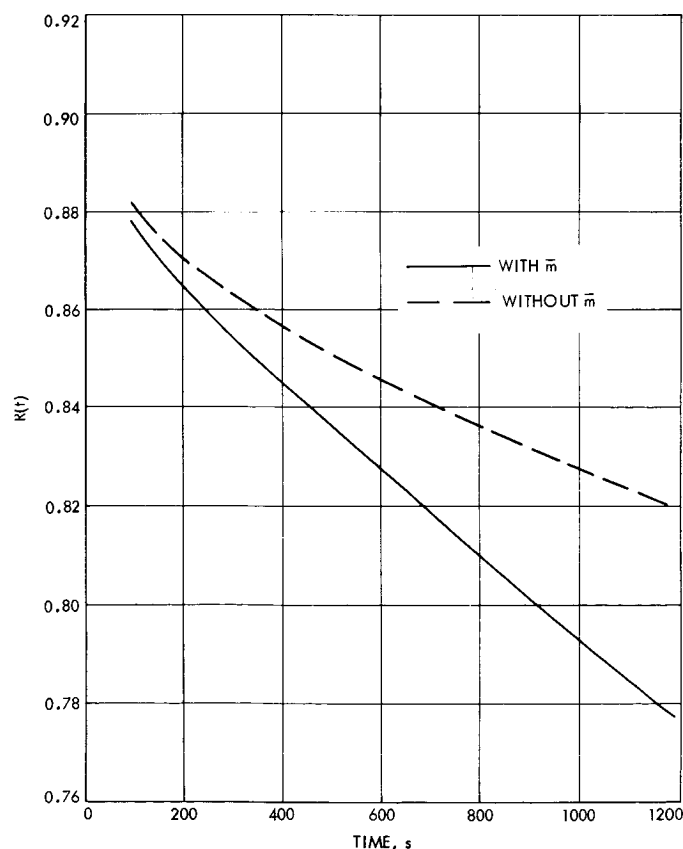


Fig. 15. Stress relaxation response

to make both effects sizable in the region of the calculation. The parameters chosen were:  $\tau = 1.0$  s,  $m = 1.0$ ,  $b = 2.0$ ,  $k = 1.0 \times 10^{-4} \text{s}^{-1}$  and  $a_T = 1.0$ .

With these parameters, 10% of the network chain would have experienced scission. After 1000 s, all approximations made in the previous discussion are valid for such extents of reactions. The dashed line in Fig. 15 depicts the predicted relaxation response of a relaxing-reacting polymer characterized by the parameters mentioned. The relaxation function  $R(t)$  is the stress normalized to an arbitrary time, in this case to 7.15 s. This predicted behavior is contrasted with the solid line calculated without the  $\bar{m}$  shift.

### 3. Summary

This technique purports to calculate the stress relaxation or creep behavior of polymers experiencing chain scission. The stress relaxation or creep master curve for a polymer at a reference crosslink density must be known along with information concerning Plazek's parameter to allow prediction of the response at other crosslink densities. These data, in conjunction with information on the kinetics of scission reactions, represent the complete input required for the calculation.

### Erratum

In Space Programs Summary 37-58, Vol. III, p. 200, Eq. (12) should have been expressed as follows:

$$\frac{d\bar{m}}{dv} = \frac{t + \bar{m}}{m\nu} \left[ \left( \frac{t + \bar{m}}{\tau a_v} \right)^m + 1 - mb \right]$$

## G. Long Term Aging of Elastomers: Chemical Stress Relaxation and Other Studies of Peroxide-Cured Styrene-Butadiene Rubber, R. Rakutis, S. H. Kalfayan, and R. H. Silver

### 1. Introduction

The chemical reactions that cause rubbers to age are investigated. Styrene-butadiene rubber (SBR) is, at the present, the elastomer selected for investigation. The methods used in the study include stress relaxation, swelling measurements, and infrared spectrophotometry.

At elevated temperatures, stress relaxation of an elastomer is mainly attributable to a chemical rupture of the rubber network. This rupture, as discussed in SPS 37-59,

Vol. III, pp. 194–198, is ascribed to the effect of molecular oxygen on the network. At higher temperatures, both crosslinking and scission occur simultaneously. The predominance of one process over the other depends on the type of rubber and the nature of the crosslink (Ref. 1).

The technique of continuous stress relaxation and intermittent stress relaxation measurements was proposed by Tobolsky (Ref. 1) as a method for separately measuring the effects of scission and crosslinking (Refs. 2 and 3).

Flory and Rehner (Refs. 4 and 5) have derived an expression that gives the apparent number of effective network chains  $\nu'_e$ , per  $\text{cm}^3$  of gel. This expression is in terms of the volume fraction of polymer  $v_2$  in the swollen network that contains tetrafunctional crosslink sites and is in equilibrium with pure solvent. The expression follows:

$$\nu'_e = \frac{\nu_e}{g} = \frac{1}{V_s} \frac{\ln(1 - v_2) + v_2 + \mu v_2^2}{v_2^{1/3} g^{2/3} - \frac{v_2}{2}} \quad (1)$$

where  $V_s$  is the molar volume of the solvent,  $\mu$  is the polymer-solvent interaction coefficient, and  $g$  is the gel fraction of the network. The volume fraction of gel can be calculated by the volume method (i.e., weighing the swollen gel) and by the linear method (i.e., measuring the sample length in the dry and swollen states). Equation (2) gives  $v_2$  as calculated by the swelling method, and Eq. (3) gives  $v_2$  by the linear method:

$$v_2 = \frac{w_g/\rho_g}{w_g/\rho_g + w_s/\rho_s} \quad (2)$$

$$v_2 = \left( \frac{\text{dry length}}{\text{swollen length}} \right)^3 \quad (3)$$

where  $w_g$ ,  $w_s$ ,  $\rho_g$ , and  $\rho_s$  are the weights and densities of the gel and solvent, respectively.

Inspection of the infrared spectra of oxidized (aged) SBR had shown (SPS 37-57, Vol. III, pp. 166–169) increased absorption taking place in several regions. An attempt is being made to make quantitative estimations of these changes and, if possible, to obtain kinetics data.

## 2. Experimental Method

**a. Stress relaxation measurements.** The preparation of SBR films and the stress relaxometer used in these experiments have been described in SPS 37-59, Vol. III,

pp. 194–198. Styrene-butadiene rubber, Batch 2, cross-linked with 0.3% dicumyl peroxide, was used for continuous-intermittent stress relaxation and also for the swelling measurements.

The experimental conditions for the continuous stress relaxation were the same as described in SPS 37-59, Vol. III. For the intermittent stress measurements, the film was extended 29% for 6 min and relaxed for 1 h at 100°C. The duration of the test was about 28 h. The values of current force at a particular time  $f_t$  were read directly off the semilogarithmic plot of  $f_t/f_0$  vs time.

**b. Swelling measurements.** A correction was made for the non-rubber compounding ingredients by subtracting 7% of the weight of the SBR rubber film from the starting weight of the rubber. From previous measurements, it was found that approximately 7% non-rubber ingredients were extracted by acetone. The samples were swollen with benzene at room temperature for 7 days. From the weights of the swollen and dried gel, the weight of absorbed solvent was determined by difference. The sample length was measured in both the dry and swollen states with a cathetometer. The gel fraction of the network was found to be 0.93.

**c. Infrared measurements.** A 4-mil-thick SBR film (crosslinked with 0.3% dicumyl peroxide and washed with acetone to remove the anti-oxidant and other non-rubber ingredients) was used in the infrared spectral study. A Perkin-Elmer 421 recording spectrophotometer was used for the measurements. The reference beam was balanced by an identical untreated film during all of the measurements.

The film was heated for 1-h intervals in air at 115°C, and cooled each time at  $-20^\circ\text{C}$  for 20 min to quench the reaction. The infrared measurements were taken at room temperature after each cooling interval.

## 3. Results and Discussion

**a. Crosslinking from intermittent stress relaxation.** In continuous stress relaxation, a sample is continuously maintained at a fixed elongation and the stress is measured as a function of time. In intermittent stress measurements, the sample is maintained in a relaxed condition at a suitable constant temperature. At evenly spaced intervals, it is rapidly stretched to a fixed elongation, the quasi-equilibrium stress is measured, and the sample is returned to its unstretched length. Actually, it is a periodic determination of the elastic modulus. If a crosslinking

reaction is occurring along with the scission, the network chains thus formed in the relaxed state will be indistinguishable from the original chains. These new chains will also contribute to the modulus. Therefore, intermittent relaxation measures the net result of both scission and crosslinking during network breakdown. If only scission were occurring, the continuous and the intermittent stress relaxation curves would be identical. If, on the other hand, the stress at a given time in the intermittent experiment is larger than in the continuous experiment, the difference represents the contribution of the crosslinking reaction.

It is possible to calculate the number of chains formed for the elastomer under investigation, from the equation of state for stress relaxation in a network,

$$\nu_e = \frac{f_0}{ART \left( \lambda - \frac{1}{\lambda^2} \right)} \quad (4)$$

where  $f_0$  is stress at time zero,  $A$  is the cross-sectional area of the undeformed specimen,  $\lambda$  is the extension ratio,  $R$  is the gas constant, and  $T$  is the absolute temperature.

The results of continuous and intermittent stress relaxation are given in Figs. 16 and 17. The values in Fig. 17 were obtained from Eq. (5), (a modification of Eq. 4) by the use of the values for  $f_t/f_0$  directly from the graphs.

$$\nu_e(t) = \frac{1}{RT \left( \lambda - \frac{1}{\lambda^2} \right)} \left[ \frac{\left( \frac{f_t}{f_0} \right)^i f_0^i}{A^i} - \frac{\left( \frac{f_t}{f_0} \right)^c f_0^c}{A^c} \right] \quad (5)$$

As reported previously (Ref. 1), there is a considerable amount of crosslinking taking place along with scission during air oxidation of SBR. The increase in the number of moles of chains does not appear to be linear (Fig. 17), but there is a rapid increase in the rate during the latter stages of the oxidation. It would be of interest to examine the effect of a chemical additive such as tertiary dodecyl mercaptan on the extent of crosslinking taking place. It has been reported (Ref. 5) that with natural rubber, the addition of thiophenol causes the continuous and intermittent stress relaxation curves to coincide.

**b. Swelling measurements.** The values for  $\nu_2$  and  $\nu_e$  are tabulated below:

Volume fraction of polymer	Volume method (Eq. 2)	Linear method (Eq. 3)
$\nu_2$	0.153	0.144
Number of chains formed	Swelling measurements (Eq. 1)	Stress relaxation measurements (Eq. 4)
$\nu_e$	$1.10 \times 10^{-4}$	$9.95 \times 10^{-5}$

There is reasonably good agreement between the  $\nu_e$  values obtained from the swelling measurements (Eq. 1) and those obtained from the stress-strain equation (Eq. 4). The 10% discrepancy is similar to that found previously for unextracted, peroxide-cured SBR (Ref. 6), and undoubtedly represents the contribution to the modulus of the entanglement network. The contribution of this transient network has been ignored in the preliminary analysis given here, but is the main subject of the overall investigation (SPS 37-57, Vol. III, pp. 169-172). The number of moles of chains formed during stress relaxation is apparently less than the number present at the start.

**c. Infrared measurements.** As observed previously (SPS 37-57, Vol. III, pp. 166-169), the two main regions of absorption during the oxidation of the SBR film were the  $-\text{OH}$  region at  $3450 \text{ cm}^{-1}$  and the  $\text{C}=\text{O}$  region at  $1700 \text{ cm}^{-1}$  (Fig. 18). There was also a general broad absorption in the  $1100 \text{ cm}^{-1}$  region. These results agree with several similar studies reported in the literature (Refs. 7 and 8). As seen in Fig. 19, the rate of  $-\text{OH}$  formation does not appear to be following any reaction order. It is plausible that the  $-\text{OH}$  peak, especially at the early stages of oxidation, is due to the hydroperoxides that are formed initially. In the photo-oxidation of polypropylene (Ref. 8), it was concluded that only 10% of the  $-\text{OH}$  peak was due to alcohol (as found by chemical analysis); the remaining amount was attributed to hydroperoxides. In this study, a 6% weight gain was noted in the film after the 5-h oxidation period.

Attempts to characterize the oxidation products, as well as improving methods of quantitative estimation of the chemical changes taking place in the aged rubber, will be continued.

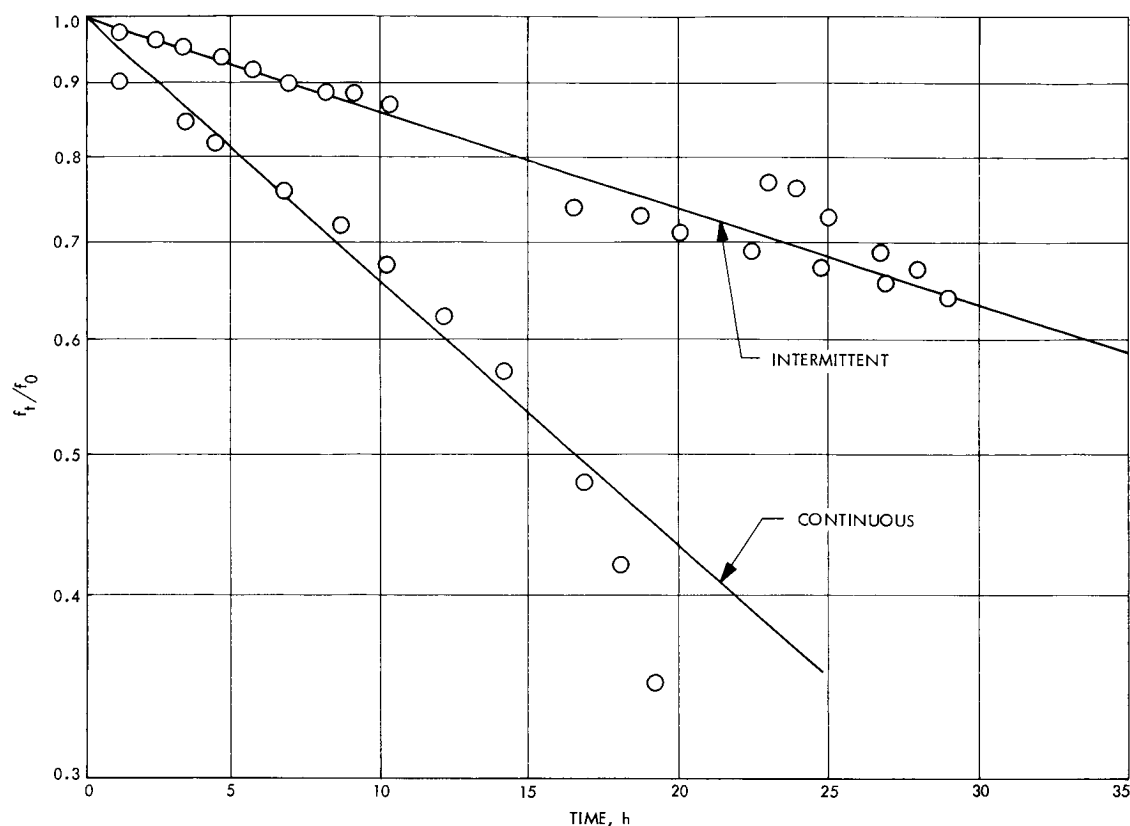


Fig. 16. Continuous and intermittent stress relaxation in SBR at 100°C in air

#### References

1. Tobolsky, A. V., *Properties and Structure of Polymers*, p. 230. John Wiley & Sons, Inc., New York, 1960.
2. Tobolsky, A. V., Prettyman, I. B., and Dillon, J. H., "Stress Relaxation of Natural and Synthetic Rubber Stocks," *J. Appl. Phys.*, Vol. 15, p. 309, 1944.
3. Tobolsky, A. V., and Andrews, R. D., "Systems Manifesting Superposed Elastic and Viscous Behavior," *J. Chem. Phys.*, Vol. 11, p. 125, 1945.
4. Flory, P. J., and Rehner, J., Jr., *J. Chem. Phys.*, Vol. II, p. 521, 1943.
5. Allen, V. R., *Structure and Degradation of Rubber Networks*, Ph.D. Dissertation. University of Akron, University Microfilms, Inc., Ann Arbor, Mich., 1960.
6. Landel, R. F., and Fedors, R. F., "A Test of the Predictability of the Properties of Filled Systems," paper presented at the AIAA meeting, Washington, June 1967.
7. Solomon, G. and Van der Schee, A. C., "Infrared Analysis of Isomerized, Vulcanized, and Oxidized Natural Rubber," *J. Polym. Sci.*, Vol. 19, p. 503, 1954.
8. Kato, U., Carlsson, D. J., and Wiles, D. M., "The Photo-Oxidation of Polypropylene: Some effects of Molecular Order," *J. Appl. Polym. Sci.*, Vol. 13, p. 1447, 1969.

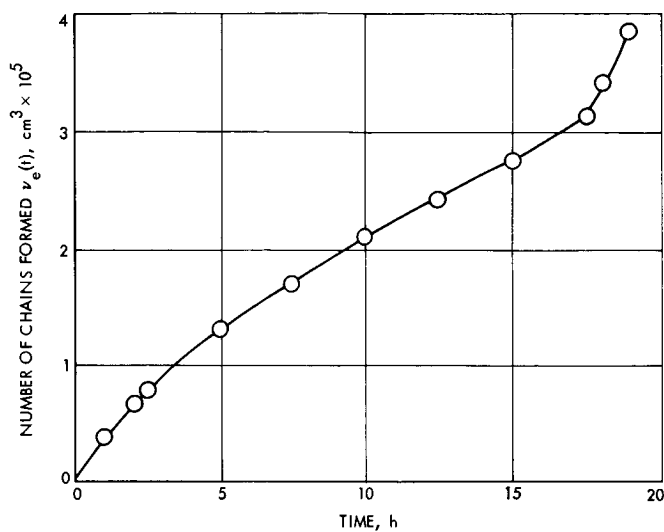


Fig. 17. Additional network chains formed

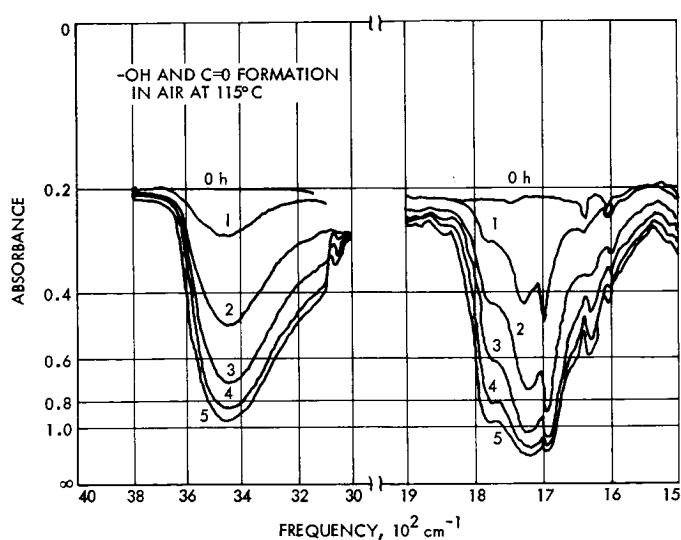


Fig. 18. Oxidation products of SBR film

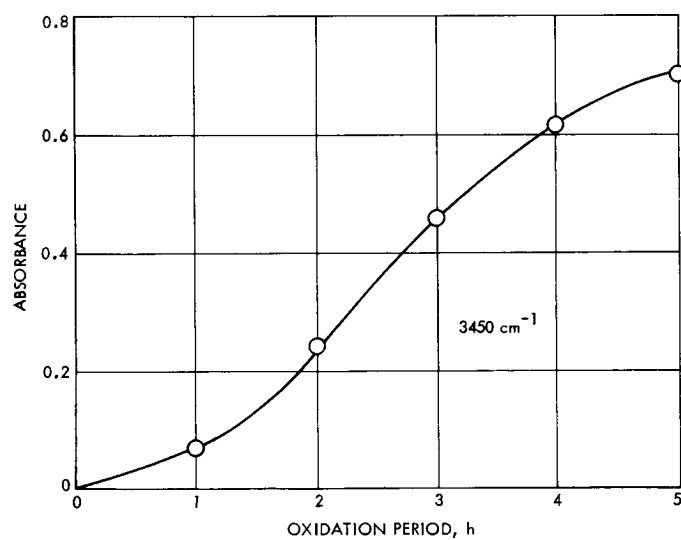


Fig. 19. Change in -OH absorption of SBR at 115°C

## XIV. Research and Advanced Concepts

### PROPULSION DIVISION

#### A. Comparison of Theoretical and Experimental Deflection of a Blown Arc, *T. K. Bose, M. B. Noel, and P. F. Massier*

##### 1. Introduction

This discussion pertains to an investigation of the influence of a transverse gas flow on the configuration of an electric arc. This work was initiated because electric discharges that become deflected in a direction normal to the applied electric field by a directed gas motion commonly occur in plasma propulsion and arc-heated devices. In plasma devices that contain segmented electrodes, an electric discharge that occurs across one pair of electrodes attaches in a small region near the trailing edge of the anode and may be deflected to the extent that the high-temperature ionized gas flows along the surfaces of downstream segments. Hence, a significant contribution to the heat transfer can occur by convection from the deflected arc to the downstream segments. This condition (which occurs in practice) of small attachment regions and ionized gas flow of the deflected arc over surfaces located downstream of electrodes differs from simplified models of plasma boundary layer flow (Ref. 1 and SPS 37-56, Vol. III, pp. 201-203). In these analyses, it is assumed that the discharge and therefore also the

current flow are perpendicular to the gas flow direction. However, the direction of the current flow in the deflected arc may be in the gas flow direction and parallel to the downstream surfaces. Furthermore, in a device in which a cool gas approaches the electrode region, the gas flow may remain cool in the central core between the electrodes for a considerable distance as can be observed in Fig. 1. This figure was obtained during a test associated with a heat transfer investigation of segmented electrodes (SPS 37-59, Vol. III, pp. 227-233). A discussion of heat transfer is beyond the scope of this article; however, a comparison will be made between predicted and experimental deflection of an arc discharge.

##### 2. Theory

One of the theoretical results of the present investigation is that for given electric field and gas flow field distributions, the dimensionless maximum deflection of the arc  $D/L$  (Fig. 2) is dependent on a dimensionless blowing parameter (Table 1)

$$B = \frac{\pi \sigma_c U_\infty \delta^2}{4 \mu_e I} \quad (1)$$

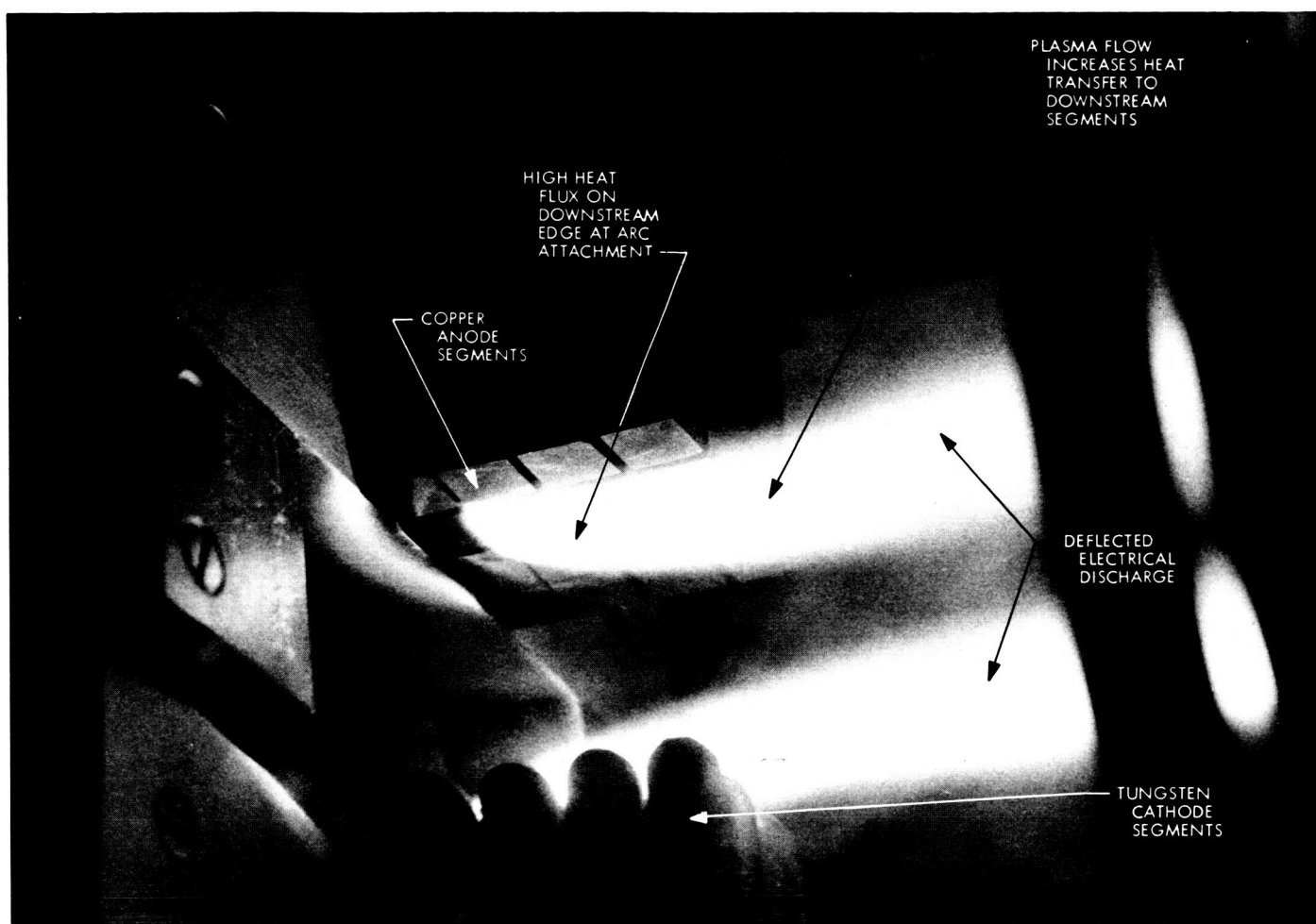


Fig. 1. Arc deflected by blowing

This result applies for the round arc that is considered here. The basic relationship, however, was obtained by considering the motion of the electrons in a two-dimensional arc that had the same average current density and the same thickness as the diameter  $\delta$  of an equivalent round arc. For comparison with the theory, all experimental arc columns are assumed to have round cross-sections of diameter  $\delta$  (Fig. 2). In the theoretical analysis, the resultant deflected motion of the electrons is considered to be caused by a combination of the electric field forces acting in the direction of the externally applied electric field and the collisional forces that result from collisions of the electrons with the heavy particles of the gas cross-flow.

In Eq. (1), the electrical conductivity  $\sigma_c$  and the mobility coefficient for the electrons  $\mu_{ec}$  are the thermophysical properties that must be evaluated to compute the blowing parameter  $B$ . These properties are dependent on the pres-

sure and the temperatures of the electrons and the heavy particles. Because it is generally difficult to evaluate the temperatures of these particles if they differ, it is assumed that the temperatures are equal. In addition, the arc diameter  $\delta$  is considered to be dependent on the arc current only. This condition is true for arcs with currents above 5 A and may be explained by considering the Elenbaas-Heller model (Ref. 2). In this model it is assumed that the external gas flows around the arc, and consequently the energy equation contains only the conduction and the joule heating terms. From the solution of the energy and the current conservation equations for this model, it was found that cylindrical arcs of different diameters have the same relative temperature distribution for a fixed power ( $I \cdot E$ ) if the radiation is neglected. Additionally, because the electric field  $E \propto (I/\delta^2)$ , the relative temperature distribution is dependent only on  $(I/\delta)^2$  (Ref. 3). The field strength  $E$  measured in nitrogen arcs for currents above 5 A shows that  $E$  is independent

**Table 1. Nomenclature**

$B$	blowing parameter
$E$	electric field
$I$	arc current
$p$	pressure
$U_{\infty}$	cross-flow velocity
$\delta$	arc diameter
$\mu_{ec}$	mobility coefficient of the electrons at the centerline of the arc
$\sigma_c$	electrical conductivity at the centerline of the arc

of the current and depends only to a small extent on the pressure (Ref. 2). Thus, if  $E$  is kept constant,  $I \propto \delta^2$  and the temperature distribution is dependent on the current only. In Ref. 2, Finkelnburg and Maecker point out that for currents above 100 A, the diameter remains approximately constant because of the self-induced magnetic compression. Thus, for a constant current flow, all thermophysical properties at the centerline of the arc, including those necessary to compute the blowing parameter  $B$ , are dependent on the pressure only. There is a very weak dependence of the electrical conductivity on the pressure, but the mobility coefficient is inversely proportional to the pressure.

From Eq. (1) it can be shown that if the current  $I$  remains constant,

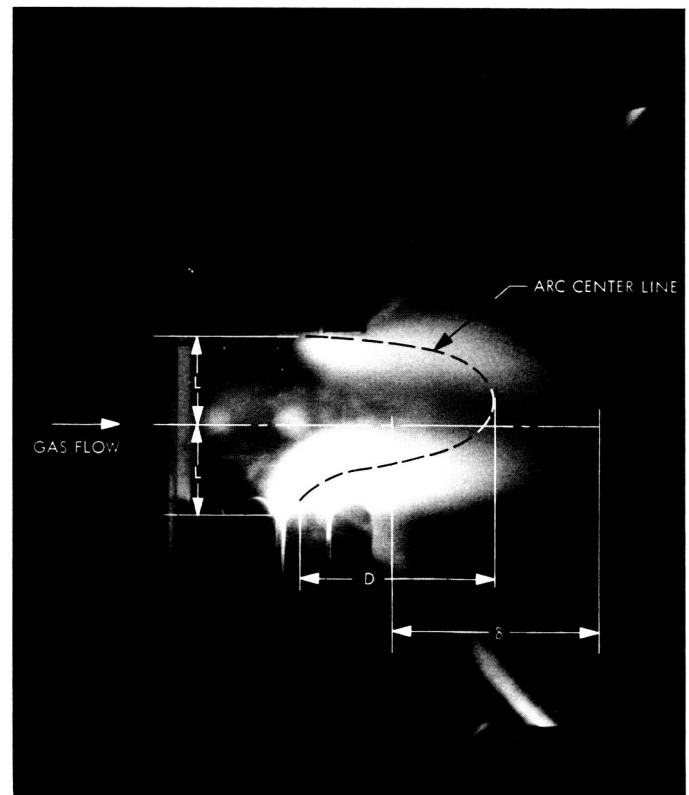
$$\frac{D}{L} = f(p U_{\infty}) \quad (2)$$

The dependence of the arc deflection on pressure and velocity was apparent from the experiments of Thiene (Ref. 4) and Baranov (Refs. 5-7). Baranov determined an empirical relation for the deflection  $D/L$  that was found to be dependent on a blowing parameter  $B' = 4U_{\infty}L/D_{amb}$ . Since the ambi-polar diffusion coefficient  $D_{amb}$  is inversely proportional to the pressure, the same relation as Eq. (2) can be obtained in this case also.

### 3. Experiments and Results

The electrode configuration used in the experiments to determine the shape of a blown arc consists of a matrix of flat, electrically and thermally insulated anode segments with an electric field applied to only one pair of the upstream segments (Fig. 1). Thus, there was no cur-

rent flow to the downstream segments. Figure 1 shows two streams of luminous plasma flow between the flat copper anode segments above and the rounded tungsten cathodes below (there are 16 pairs of electrodes). The two luminous streams of plasma are actually a single electrical discharge between one pair of the electrodes. This discharge has been deflected from a vertical plasma column into an elongated U-shape by the motion of argon from left to right; however, the view does not extend sufficiently downstream to show that the two plasma streams are actually connected. The size of each anode segment is  $0.25 \times 0.25$  in., the space between the segments are 0.040-in. wide, and the distance between the copper anodes and tungsten cathodes is 1.0 in. Tests were conducted over the following range of conditions in an argon atmosphere: current between 40 and 100 A, gas velocity between 10 and 250 ft/s, and gas pressure between 30 and 400 torr. The deflection  $D$  and the arc diameter  $\delta$  were determined from the photographs by measuring the maximum distance of the luminous region from the line connecting the centers of the electrodes and by measuring the thickness of the arc columns, respectively (Fig. 2). The accuracy in the measurement of the deflection is  $(\Delta D/D) = \pm 0.2$  and for the arc thickness



**Fig. 2. Definition of characteristic lengths for a blown arc**

is  $(\Delta\delta/\delta) = \pm 0.2$ . The test results for deflection are shown as a function of the product of velocity and pressure in Fig. 3 with the arc current as a parameter. The test results for different currents do not fall exactly on one line. This may be the result of the limited accuracy of the deflection measurements or else the magnitude of the electric field may be slightly dependent on the current. However, because of the small scatter of the data, it can be deduced that the arc thickness may be approximated by assuming it to be directly proportional to the square root of the arc current and the electric field is virtually independent of the current. Furthermore, the test results plotted in Fig. 3 show the validity of Eq. (2).

To evaluate the blowing parameter  $B$ , it is necessary to know the temperature of the arc. This temperature was not measured during these experiments but was estimated based on previous experiments.<sup>1</sup> A maximum temperature of  $15,000^\circ\text{K}$  was used to evaluate  $\sigma_c$  and  $\mu_{ec}$ . Theoretical results of the deflection based on a model of an arc between two point electrodes and a uniform cross flow velocity distribution are shown in Fig. 4 as a function of the given value of the blowing parameter  $B$ . From

<sup>1</sup>Bosnjakovic, F., and Associates, unpublished results of experiments, 1960-1965, Technical Univ., Stuttgart, Germany.

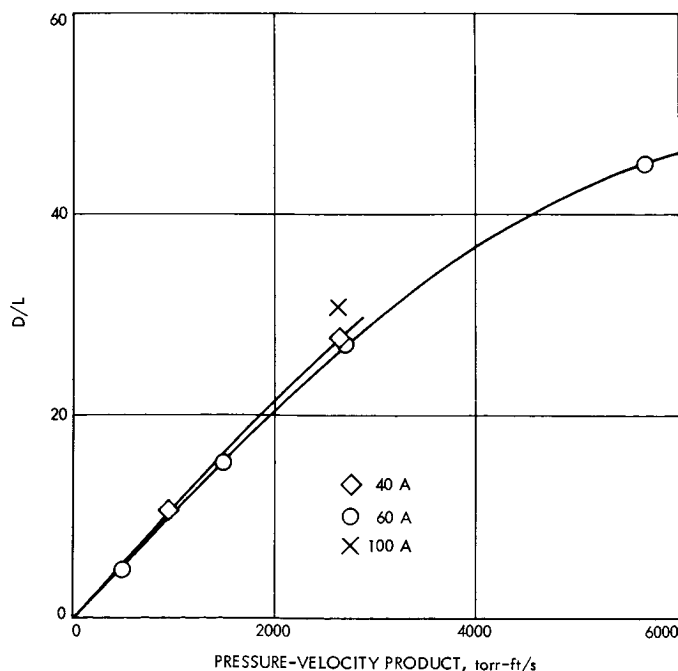


Fig. 3. Dimensionless arc deflection vs the product of the pressure and the cross-flow velocity

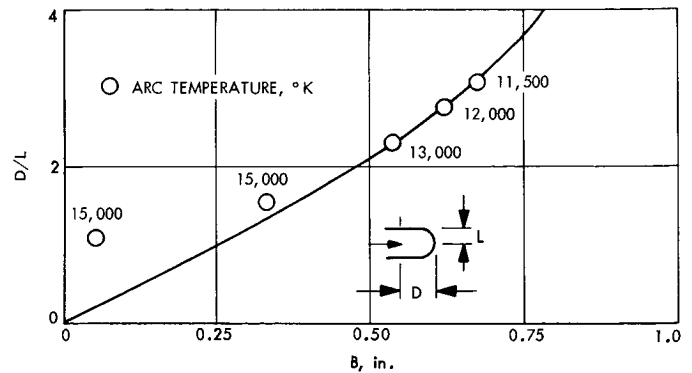


Fig. 4. Comparison of experimental and theoretical deflections

the experimental deflection data and by comparing with the theoretical results, an experimental  $B$  was obtained from which the value of temperature on the centerline of the arc was determined. In Fig. 4, the temperatures obtained in this manner from several experiments are given with a maximum limit of the temperature at  $15,000^\circ\text{K}$ . Though the method is not very accurate, a definite tendency is shown that as a result of blowing the maximum temperature in the arc is reduced.

#### 4. Concluding Remarks

The maximum deflection of an arc at various flow conditions was compared with theoretical predictions that are based on electrical and collisional forces between the electrons and the heavy particles in the arc region. It was concluded that:

- (1) The arc deflection is dependent on the product of the pressure and the cross-flow velocity.
- (2) The arc thickness is approximately proportional to the square root of the arc current in the range between 40 and 100 A.
- (3) At a given current the temperature at the centerline of the arc is reduced when blowing occurs.

#### References

1. Sherman, A., and Reshotko, E., "Non-equilibrium Boundary Layer Along an Insulator Wall," *AIAA J.*, Vol. 7, No. 4, pp. 610-615, Apr. 1969.
2. Finkelburg, W., and Maecker, H., "Elektrische Bogen und Thermisches Plasma," *Handbuch der Physik*, Volume 22, pp. 386-402. Springer-Verlag, Berlin, 1956.
3. Bose, T., Kalra, S. P., and Murthy, S. N. B., *Design Charts for Nitrogen Plasma Tube Arc Heaters*, Report TN-PR-2-660. National Aeronautical Laboratory, Bangalore, India, Feb. 1966.

4. Thieme, P. G., "Convective Flexure of a Plasma Conductor," *Phys. Fluids*, Vol. 6, No. 9, pp. 1319-1324, Sept. 1963.
5. Baranov, V. Yu, and Vasileva, I. A., "An Electric Arc in a Flow of Argon," *High Temperature*, Vol. 2, pp. 609-615, 1964.
6. Baranov, V. Yu, and Vasileva, I. A., "Study of the Nonisothermal Plasma on an Arc in a Flow of Argon," *High Temperature*, Vol. 3, No. 1, pp. 1-11, 1965.
7. Baranov, V. Yu, "Some Effects Noticed When Studying an Electric Arc in a Gas Flow," *High Temperature*, Vol. 4, No. 5, pp. 585-587, 1966.

## B. Liquid-Metal MHD Power Conversion,

D. G. Elliott, D. J. Cerini, L. G. Hays, and  
D. W. Bogdanoff

### 1. Introduction

Liquid-metal magnetohydrodynamic (MHD) power conversion is being investigated as a power source for nuclear-electric propulsion. A liquid-metal MHD system has no moving mechanical parts and operates at heat-source temperatures between 1200 and 1400°K. Thus, the system has the potential of high reliability and long life-time with the use of readily available containment materials such as Nb-1%Zr. Testing of the NaK-nitrogen conversion system has begun, assembly of the Cs-Li loop is continuing, and further experiments have been conducted on low-friction separation with impinging nozzles.

### 2. NaK-Nitrogen Conversion System

The NaK-nitrogen conversion system was described in SPS 37-51, Vol. III, pp. 120-124. The system is designed to demonstrate integrated hydraulic and electrical operation of a liquid-metal MHD power source. The unit employs a circular nozzle, a flat surface separator, and an induction generator. Diffusers upstream and downstream of the generator each contain seven electrically insulating vanes, which, in conjunction with compensating poles, reduce generator end losses. Startup is achieved by setting a nitrogen flow rate of 2.5 lb/s through the nozzle and injecting the NaK inventory at 110 lb/s from a 30-gal start tank. The start-tank flow is shut off by check valves when closed loop NaK flow is established through the nozzle, separator, generator, diffuser, and return lines. A 200-gal tank supplies 0.5-5.0 lb/s NaK to compensate for the inventory transferred to a droplet separator by the nitrogen exhausting to atmosphere.

On the first two runs, abnormally high makeup flow rates were encountered and generator excitation was not attempted. A post-run inspection revealed a partial

blockage of the generator flow channel by one of the fiberglass coil-slot covers. Removal of the blockage permitted proper starting in runs 3-5 at a nozzle inlet pressure of 105 psia and NaK flow rate of 110 lb/s. A smooth transition to a nozzle pressure of 150 psia and NaK flow rate of 120 lb/s was accomplished by additional NaK injection from the supply tank and increased nitrogen flow.

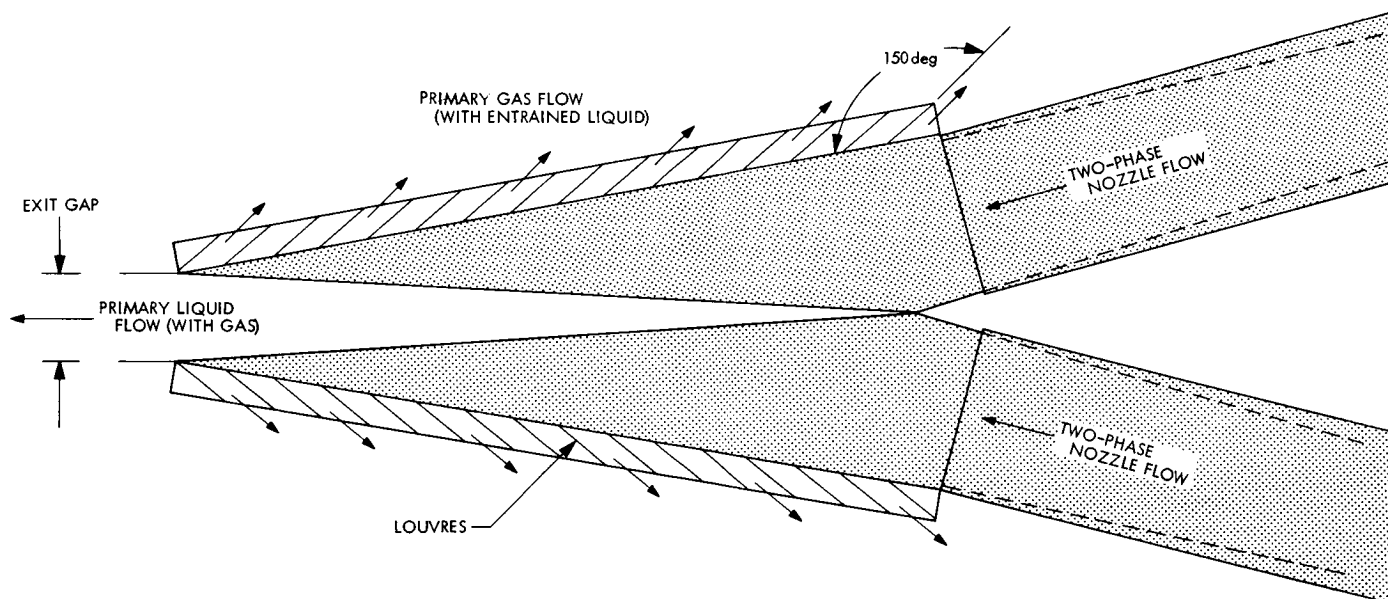
The coil-slot cover loss had resulted in a NaK leak into the generator housing. Flushes with kerosene, alcohol, and ammonia removed the NaK shorts, and empty-channel generator excitation verified generator readiness. On runs 6 and 7, power generation was attempted at the starting condition of 105 psia nozzle inlet pressure by excitation of one of the eight generator phases, but the remaining seven phases failed to excite. Additional NaK shorting was found after the runs and ammonia flushing was again required.

On run 8, external excitation was applied to all phases, and the effect of starting the NaK flow was found to be a doubling of the input power to the generator. This indicated that the NaK velocity was below the wave speed and the NaK was being pumped. This NaK shorting prevented further runs with higher NaK velocity and the system has been disassembled for repair of the channel wall.

### 3. Impinging Nozzle Tests

Investigation of low-friction, liquid-phase concentration and separation is continuing. The impingement of the two phase flows from two nozzles has produced liquid concentrations about three times higher than that at the nozzle exit. This means that a surface separator to perform the final liquid-gas separation would require only one-third the area (and hence friction loss) of a full surface separator.

Liquid loss with the gas leaving the impinging nozzle separator is a greater problem than in a surface separator. Reduction of this loss with louvers inclined at an angle to the flow was accomplished with the configuration shown in Fig. 5. Gas flow leaving the nozzles is turned through 150 deg by the louvers. The high turning angle results in separation of the gas from small drops of entrained liquid. The louvers are inclined at 30 deg to the main flow direction to minimize momentum loss of large drops. Figure 6 is a photograph of the test section during operation with the nozzles inclined 10 deg to the axis.



**Fig. 5. Louvred test section**

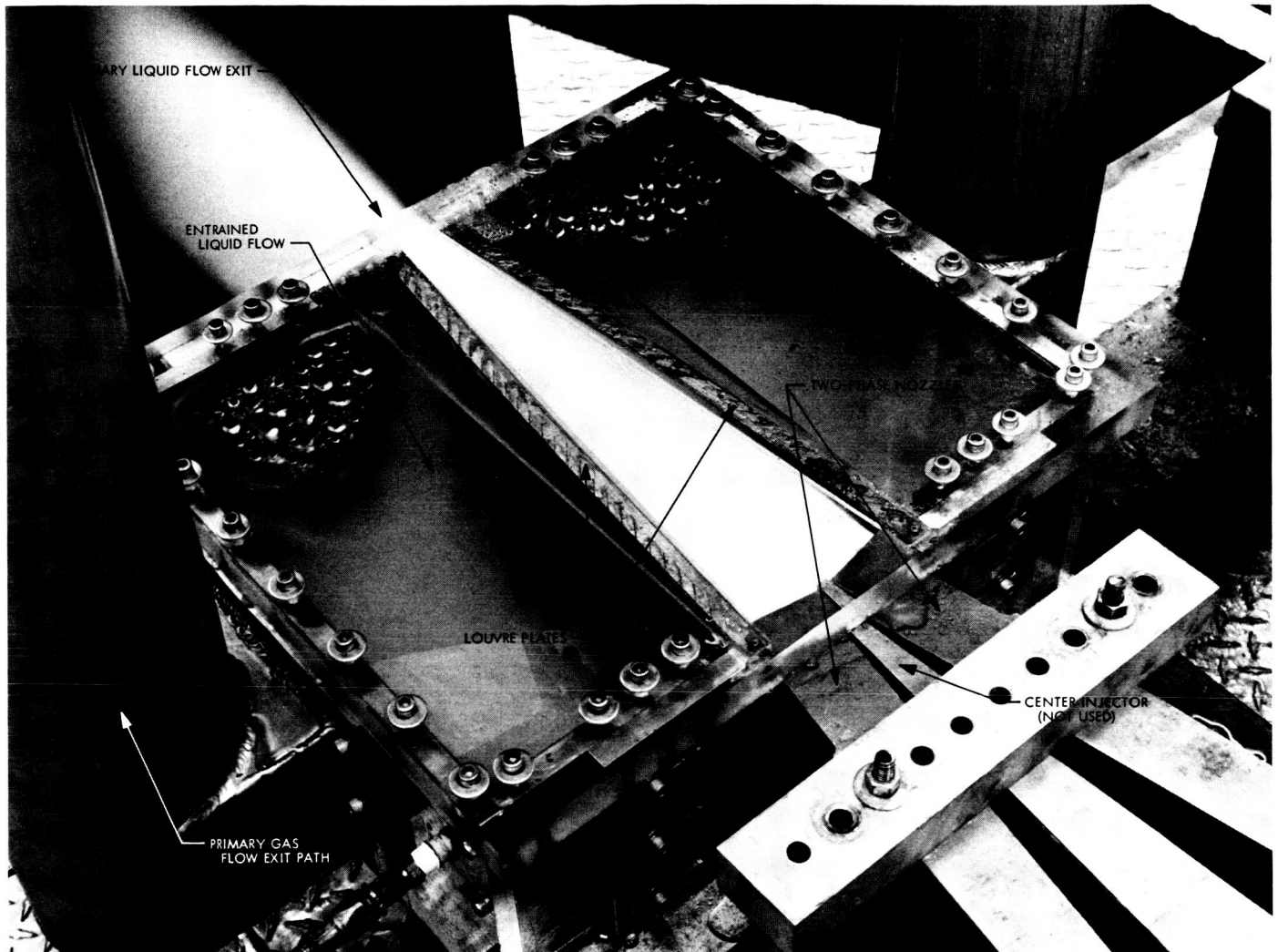
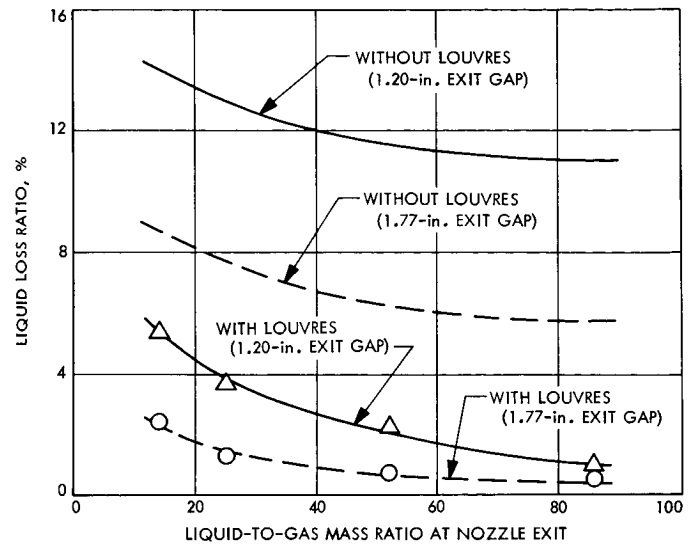


Fig. 6. Louvred test section during operation

Initial tests have shown the louvre configuration to be very effective in the reduction of liquid loss. Figure 7 gives the liquid loss ratio for two louvre exit gap settings as a function of the nozzle liquid-to-gas mass ratio at the nozzle exit. At a liquid concentration ratio of 2.7 (1.20-in. exit gap), the liquid loss range was from approximately 5% at the lowest mass ratio to 1% at the highest. These values can be compared to liquid losses of 14–11% which were obtained for the same mass ratios and exit gap without the louvres. At a concentration ratio of 1.8 (1.77-in. exit gap), the liquid loss ranged from 2.5 to 0.5% compared with a range of 9–6% for the impinging nozzles without louvres. Preliminary data indicate only a small velocity loss with the louvres compared to the flow without louvres. Thus, the additional liquid retained should be available for power generation whereas the liquid momentum is essentially lost in the impinging nozzles without louvres. Refinement of velocity results and tests of additional exit gap settings are in progress.



**Fig. 7. Liquid loss ratio for 10-deg impinging nozzles with and without louvres**

## XV. Liquid Propulsion

### PROPULSION DIVISION

#### **A. Material Compatibility, O. F. Keller and L. R. Toth**

A material compatibility program is under way to evaluate the effects of long-term storage of earth-storable propellants with selected materials of construction for spacecraft propulsion systems. The current program, Phase II, was described in SPS 37-60, Vol. III, pp. 211-215. In these immersion tests, all aspects of the total system are considered and include propellant, material, test fixture, test container, instrumentation, test environment, and facility.

The number of samples including reference units selected for testing are distributed between fuel and oxidizer propellants as follows:

Hydrazine ( $N_2H_4$ ), 392 units

Neat monomethylhydrazine ( $CH_3NH_2$ ), 112 units

Hydrazine/hydrazine nitrate mix ( $N_2H_4/N_2H_5NO_3$ ), 122 units

Inhibited nitrogen tetroxide ( $N_2O_4$ ), 405 units

To date, approximately 500 specimen/capsules have been delivered to the Edwards Test Station. From this quantity, a small number of hydrazine units were installed in the fuel test system for calibration and checkout purposes. These are still being monitored at an ambient temperature of 110°F, and capsule internal pressures are automatically being recorded on data cards. The oxidizer test fixture electrical wiring has been completed, and the test system is now ready for installation of calibration units. The balance of the delivered units have been temporarily placed in cold storage pending completion of the respective checkout phases.

The automatic data processing begins with the data card output of the automatic data acquisition system. A computer program is being written to read the data cards and refine the raw data. The final output of the computer program will consist of graphical plots of specimen/capsule internal pressure versus time. An additional function of the automatic data processing system will be to identify any specimen/capsule that exhibits a rapidly rising internal pressure, as part of an alarm system

to prevent specimen/capsule overpressurization. Approximately 2 h are required for a complete scan of all data channels. Readings are taken on a daily basis or on command.

Three possible data shifts within the automatic data acquisition system have been identified as follows:

(1) Capsule/transducer sensitivity changes that are caused by variations in calibration techniques between the capsule/transducer manufacturer and the test facility at the Edwards Test Station.

(2) Capsule/transducer zero shift changes that are attributed to epoxy-bond shrinkage between the strain-gage-type transducer and the glass capsule.

(3) Input-voltage changes that are caused by variations of input-voltage drop to various locations within the test bay, resulting in variations of output voltage from the capsule/transducer.

The above variations are being factored into the data processing.

## XVI. Systems Analysis Research

### MISSION ANALYSIS DIVISION

#### A. Advances in the Numerical Integration of the Motion of the Moon, D. B. Holdridge,

K. K. Garthwaite, and J. D. Mulholland

Since the discussion presented in SPS 37-57, Vol. II, pp. 51-53, the burden of the numerical integration of the motion of the moon has been shifted to a system of 1108 programs called LPLINT (Lunar and PLANetary INTegration), which simultaneously integrates both the equations of motion of the moon and the nine major bodies of the solar system (the earth-moon barycenter is integrated directly instead of only the earth). LPLINT is a natural generalization of PLINT (SPS 37-49, Vol. III, pp. 1-14) and was first developed on the IBM 7094 computer.

Comparing LPLINT to PLOD (PLANetary Orbit Determination) (Ref. 1), it is to be noted that the lunar model has been extended to include the relativistic equations of motion using the isotropic metric and the effect of the lunar oblateness. The only effect that has yet to be considered is that of tides; a suitable model will be necessary before a complete comparison with observations will be possible.

The lunar oblateness model  $\Phi_\epsilon$  is represented by the following potential:

$$\Phi_\epsilon = \frac{\mu_\epsilon a_\epsilon^2}{R^3} \left[ \frac{1}{2} J_{2\epsilon} \left( 1 - 3 \frac{z^2}{R^2} \right) + 3 \left( 1 - \frac{z^2}{R^2} \right) \left( C_{22} \frac{x^2 - y^2}{x^2 + y^2} + S_{22} \frac{2xy}{x^2 + y^2} \right) \right]$$

where  $\mu_\epsilon$  is the gravitational constant of the moon,  $a_\epsilon$  is

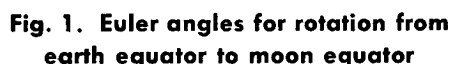
the radius of the moon,  $\mathbf{R} = (X, Y, Z)$  is the position vector of the earth relative to the moon referred to the 1950.0 equatorial system, and  $\mathbf{r} = (x, y, z)$  is the position vector of the earth referred to the equatorial system of the moon. For purposes of computational economy, and regarding the small size of the perturbing acceleration, nutations and librations can be ignored and the entire computation can be carried out in single-precision (the perturbation affects the tenth figure of the acceleration).

The coordinate transformation is defined as follows:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = (N_{ij}) (A_{ij}) \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

where  $A$  is the precession matrix and  $N$ , the Euler angle rotation matrix, is computed according to the following rules (Ref. 2).

$$\begin{aligned} N_{11} &= \cos \Lambda \cos \Omega' - \sin \Lambda \sin \Omega' \cos i \\ N_{12} &= \cos \Lambda \sin \Omega' + \sin \Lambda \cos \Omega' \cos i \\ N_{13} &= \sin \Lambda \sin i \\ N_{21} &= -\sin \Lambda \cos \Omega' - \cos \Lambda \sin \Omega' \cos i \\ N_{22} &= -\sin \Lambda \sin \Omega' + \cos \Lambda \cos \Omega' \cos i \\ N_{23} &= \cos \Lambda \sin i \\ N_{31} &= \sin \Omega' \sin i \\ N_{32} &= -\cos \Omega' \sin i \\ N_{33} &= \cos i \end{aligned}$$



253

integration method in that it can process data in real time and requires less memory space because it is unnecessary to store the filtering solution.

## 2. Sequentially Correlated Process

According to Ref. 1, a sequentially correlated process is defined as a process such that its normalized correlation matrix  $L$  has the property

$$L(t, \tau) = L(t, s) L(s, \tau) \quad t \geq s \geq \tau \quad (1)$$

Suppose a vector  $x$  composed of the state  $x_1$  to be estimated and data noise  $x_2$  satisfy the following equations:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (2)$$

$$\frac{dx(t)}{dt} = E(t)x(t) + G(t)w(t) \quad (3)$$

where  $w(t)$  is white gaussian noise having zero mean and the power spectral density  $Q(t)$ . Also assume that the data  $y$  is expressed as

$$y = H_1 x_1 + x_2 = Hx \quad (4)$$

where

$$H = [H_1, I] \quad (5)$$

Namely,  $y$  is a perfect measurement and is not contaminated by white noise directly.

The solution of Eq. (1) becomes

$$x(t) = U(t, 0)x(0) + \int_0^t U(t, s)G(s)w(s)ds \quad (6)$$

where  $U(t, s)$  is defined by

$$\left. \begin{aligned} \frac{d}{dt} U(t, s) &= E(t)U(t, s), & t \geq s \\ U(s, s) &= I \end{aligned} \right\} \quad (7)$$

and  $I$  is an identity matrix.

Then it can be easily proved that the vector  $x$  is a sequentially correlated process because its normalized correlation matrix  $L$  satisfies Eq. (1). The filtering solutions

for a process such as  $x(t)$  were derived by Pfeiffer, based on the Gauss-Markov theorem (Ref. 1).

## 3. Derivation of the Forward Smoothing Equation

The basic technique adopted for deriving the forward smoothing formula for a sequentially correlated process is essentially identical to the one employed for the white data noise case (Ref. 2). Namely, paying attention to Eq. (6), the optimal estimate of the initial condition  $x(0)$  is derived by simultaneously estimating  $x(t)$  and  $p(t)$  (where  $p(t)$  is the accumulated contribution to  $x(t)$  from the process noise) through the filtering technique.

For this purpose, let

$$p(t) = \int_0^t U(t, s)G(s)w(s)ds \quad (8)$$

then

$$\frac{dp(t)}{dt} = E(t)p(t) + G(t)w(t) \quad (9)$$

From Eq. (6),

$$x(0) = U^{-1}(t, 0)(x(t) - p(t)) \quad (10)$$

Then the optimal smoothing estimate of  $x(0)$  having data

$$Y(t) = \{y(\tau); t \geq \tau \geq 0\}$$

is denoted  $x^*(0/t)$ , and it is given by

$$x^*(0/t) = U^{-1}(t, 0)(x^*(t) - p^*(t)) \quad (11)$$

In order to derive these filtering solutions,  $x^*(t)$  and  $p^*(t)$ , an augmented state vector  $\eta(t)$  is introduced.

$$\eta(t) = \begin{bmatrix} x(t) \\ p(t) \end{bmatrix} \quad (12)$$

Then the new process equation and observation equation are described using  $\eta$ .

$$\frac{d\eta(t)}{dt} = \tilde{E}(t)\eta(t) + \tilde{G}(t)w(t) \quad (13)$$

$$y(t) = \tilde{H}(t)\eta(t) \quad (14)$$

where

$$\tilde{E}(t) = \begin{bmatrix} E(t) & 0 \\ 0 & E(t) \end{bmatrix} \quad (15)$$

$$\tilde{G}(t) = \begin{bmatrix} G(t) \\ G(t) \end{bmatrix} \quad (16)$$

and

$$\tilde{H}(t) = [H(t), 0] \quad (17)$$

Following the same discussion as in *Subsection 2*, we find that  $\eta(t)$  is a sequentially correlated process. Therefore, the filtering technique of Ref. 1 can be applied and the solution is derived as follows:

$$\frac{d\eta^*(t)}{dt} = \tilde{B}\eta^* + \tilde{K}\dot{y} \quad (18)$$

$$\frac{d\tilde{P}(t)}{dt} = \tilde{F}\tilde{P} + \tilde{P}\tilde{F}' - \tilde{P}\tilde{M}\tilde{P} + \tilde{N} \quad (19)$$

where  $\tilde{P}(t)$  is the error covariance associated with  $\eta^*(t)$ .

The matrices  $\tilde{F}$ ,  $\tilde{M}$ ,  $\tilde{N}$ ,  $\tilde{B}$ , and  $\tilde{W}$  in the above equations are identified as follows:

$$\tilde{S} = \tilde{G}\tilde{Q}\tilde{G}' \quad (20)$$

$$\tilde{D} = \tilde{H}\tilde{S}\tilde{H}' = HSH' = D \quad (21)$$

$$\tilde{A} = \dot{\tilde{H}} + \tilde{H}\tilde{E} = [A, 0] = [\dot{H} + HE, 0] \quad (22)$$

$$\tilde{F} = \tilde{E} - \tilde{S}\tilde{H}'D^{-1}\tilde{A} \quad (23)$$

$$\tilde{M} = \tilde{A}'D^{-1}\tilde{A} \quad (24)$$

$$\tilde{N} = \tilde{S} - \tilde{S}\tilde{H}'D^{-1}\tilde{H}\tilde{S} \quad (25)$$

$$\tilde{B} = \tilde{F} - \tilde{P}\tilde{M} \quad (26)$$

$$\tilde{W} = [\tilde{A}\tilde{P} + \tilde{H}\tilde{S}]'D^{-1} \quad (27)$$

The initial conditions are different from the *a priori* values because of the initial jump due to the perfect measurement  $y(0)$ .

$$\eta^*(0^+) = \tilde{P}(0)\tilde{H}'(0)(\tilde{H}\tilde{P}\tilde{H}'(0))^{-1}y(0) \quad (28)$$

$$\tilde{P}(0^+) = \tilde{P}(0) - \tilde{P}(0)\tilde{H}'(0)(\tilde{H}\tilde{P}\tilde{H}'(0))^{-1}\tilde{H}(0)P(0) \quad (29)$$

Thus, solving Eqs. (18) and (19) provides us with the optimal estimates  $x^*(t)$  and  $p^*(t)$ , which can be used to find the initial condition estimate  $x^*(0/t)$  through the relation of Eq. (10).

However, rather than dealing with matrices and vectors of augmented size in Eqs. (18) and (19), it is possible to derive a simplified solution in the original dimension.

The derivation is in parallel with that employed in Ref. 2. When the quantities in the original dimension are simply described by removing the tildes from the top of the quantities in Eqs. (20-29), the initial condition estimate is derived as

$$x^*(0/t) = x^*(0^+) + P(0^+) \int_0^t V'(s, 0) A' D^{-1} (\dot{y}(s) - Ax^*(s)) ds \quad (30)$$

where  $V(t, s)$  is defined by

$$\left. \begin{aligned} \frac{d}{dt} V(t, s) &= B(t) V(t, s) \\ V(s, s) &= I \end{aligned} \right\} \quad (31)$$

Also, the covariance of this initial-condition estimate is denoted  $P(0/t)$  and it is obtained as

$$P(0/t) = P(0^+) - P(0^+) \int_0^t V'(s, 0) M(s) V(s, 0) ds P(0^+) \quad (32)$$

This initial condition estimation technique can be extended to the general smoothing problem, namely the problem of finding the optimal estimate at time  $t$  having the information from time 0 to  $T$  where  $t$  falls between 0 and  $T$  (i.e.  $0 < t < T$ ). Since the optimal estimate at time  $t$  having the information for  $(0, t)$  is obtained as a solution of filtering problems, the smoothing problem reduces to finding an optimal initial condition estimate (at time  $t$ ) using the remaining information for the interval  $(t, T)$  by means of the previously developed technique.

The initial condition in Eqs. (30) and (32) are replaced by the optimal filtering estimate and its error covariance.

Let the optimal smoothing estimate at time  $t$  given the information  $(0, T)$ , where  $0 \leq t \leq T$ , be denoted  $x^*(t/T)$ . Then, from Eq. (30), it reduces to the following equation:

$$x^*(t/T) = x^*(t) + P(t) \int_t^T V'(s, t) A' D^{-1} (\dot{y}(s) - Ax^*(s)) ds \quad (33)$$

Also, the error covariance of this "smoother" denoted  $P(t/T)$ , is derived from Eq. (32).

$$P(t/T) = P(t) - P(t) \int_t^T V'(s, t) M(s) V(s, t) ds P(t) \quad (34)$$

Notice that both Eqs. (33) and (34) employ only the forward integration in time.

#### 4. Reduced Filtering and Smoothing Equation

The filtering estimator  $x^*(t)$  provides not only the optimal estimate  $x_1^*(t)$  of the parameters with which we are concerned, but also the optimal estimate  $x_2^*(t)$  of correlated data noise. Since, in most cases, we are not interested in the estimates of data noise, it is desirable to remove them from the estimator. Such reduction of the dimension of matrices and vectors makes it possible to save some computing time as well as memory space.

Due to the fact that we have perfect measurements as in Eq. (4), the following relation holds.

$$y(t) = Hx^*(t) = H_1 x_1^*(t) + x_2^*(t) \quad (35)$$

From this, it can be proved that

$$HP(t) = 0 \quad (36)$$

Using these relations, the filtering solution of Bryson and Johansen (Ref. 3) in the reduced dimension can be derived from the results of Pfeiffer (Ref. 1).

The reduced forms of the forward smoothing equation are derived following similar steps taken in Subsection 3 and utilizing the relation of Eq. (35) as well as Eq. (36). The forward smoother thus obtained is denoted  $x_1^*(t/T)$  and its error covariance is  $P_{11}(t/T)$ . Subscripts 1 and 2 indicate the portion corresponding to the state  $x_1$  and data noise  $x_2$ , respectively.

$$x_1^*(t/T) = x_1^*(t) + P_{11}(t) \int_t^T V'_{11}(s, t) H'_1 D^{-1} (\dot{y}(s) - E_{22} y(s) - \bar{A}_1 x_1^*(s)) ds \quad (37)$$

$$P_{11}(t/T) = P_{11}(t) - P_{11}(t) \int_t^T \bar{V}'_{11}(s, t) \bar{M}_{11}(s) \bar{V}_{11}(s, t) ds P_{11}(t) \quad (38)$$

where  $\bar{V}_{11}(t, s)$  is defined by

$$\left. \begin{aligned} \frac{d\bar{V}_{11}(t, s)}{dt} &= \bar{B}_{11}(t) \bar{V}_{11}(t, s) \\ \bar{V}_{11}(s, s) &= I \quad t \geq s \geq 0 \end{aligned} \right\} \quad (39)$$

and  $\bar{B}_{11}$ ,  $\bar{A}_1$  and  $\bar{M}_{11}$  are given by

$$\bar{B}_{11} = F_{11} - F_{12} H'_1 - P_{11} \bar{M}_{11} \quad (40)$$

$$\bar{A}_1 = \dot{H}_1 + H_1 E_{11} - E_{22} H_1 \quad (41)$$

$$\bar{M}_{11} = \bar{A}'_1 D^{-1} \bar{A}_1 \quad (42)$$

The forward smoothing technique developed in this article can be applied to the tracking problem in low-thrust missions or for spacecraft planetary approach optical measurements, when data noise is sequentially correlated.

### References

1. Pfeiffer, C. G., *Continuous Estimation of Sequentially Correlated Random Variables*, Technical Memorandum 32-524. Jet Propulsion Laboratory, Pasadena, Calif., Oct. 30, 1963.
2. Nishimura, T., "A New Approach to Estimation of Initial Conditions and Smoothing Problems", *IEEE Trans. Aerospace Electron Sys.*, Vol. AES-5, Sept. 1969.
3. Bryson, A. E., and Johansen, D. E., "Linear Filtering for Time-Varying Systems using Measurements Containing Colored Noise", *IEEE Trans. Auto. Cont.*, Vol. AC-10, pp. 4-10, Jan. 1965.

## C. The Calculation of Probabilities of Mission Failure for Solar-Electric Propulsion Systems Utilizing Varying Numbers of Thrusters, W. Kizner

### 1. Introduction

For the proposed Jupiter fly-by mission,<sup>1</sup> the engine will consist of an array of five available thrusters configured and located as shown in Fig. 2. Initially, four thrusters are used when maximum power is available (due to the proximity of the spacecraft to the sun). After about 127 days, each of the four operating thrusters is working at 75% throttle. At this time, one thruster is shut down and three are brought back to full power. At about 199 and 326 days, the number of operating thrusters is reduced by one; and at about 470 days, the last thruster is shut down.

In modes 1, 2, and 3, where 4, 3, and 2 thrusters, respectively, are employed, it is required that either the pair 1 and 3 or the pair 2 and 4 (or both) be employed to enable the thrusters to provide control about the roll axis (Fig. 2). This is accomplished by gimbaling these pairs of thrusters. Yaw and pitch control is provided by moving the thruster array in a horizontal or vertical direction. During mode 4, when only one thruster is operating, attitude control is provided by a cold gas system.

The problem dealt with here is to calculate the probability of being able to use functioning thrusters consistent

with the requirements for attitude control for each of the four modes. The law of failure that has been chosen for this study is the exponential: the probability that any thruster is operating at time  $t$  is  $e^{-\alpha t}$ ,  $\alpha$  being a positive constant. Also, it is assumed that a failure in one thruster is independent of a failure in any other. The exponential law implies that the probability that a thruster working from  $t_0$  to  $t_1$  will fail in the interval  $(t_1, t_2)$ ,  $t_2 > t_1$ , is only dependent on the difference  $t_2 - t_1$ , and not on  $t_1 - t_0$ .

This law is often used as an approximation where it is thought that random failures, as opposed to wear, are the primary cause of failure.<sup>2</sup> For the present application, where the intended operating life of a thruster is much longer than the time of operation, this exponential law seems justified. That is, the results given here would be unaffected if the true law of failure departed from the exponential law for larger values of the time than the scheduled time of operation of the engine.

There is an interesting interpretation of the constant  $\alpha$ . It can easily be shown that the expected life of a thruster whose reliability is given by the exponential law is  $1/\alpha$ . Thus, if life tests were performed on a large number of thrusters governed by this law, an estimate of  $1/\alpha$

<sup>2</sup>*Electric Propulsion System Reliability Analysis Final Report*, Hughes Research Laboratories, Malibu, Calif. (to be published).

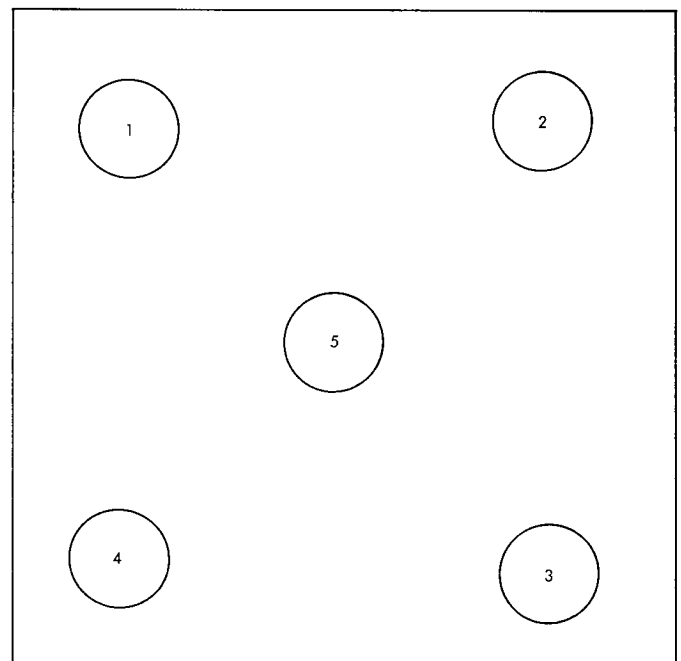


Fig. 2. Thruster configuration and location

<sup>1</sup>1975 Jupiter Flyby Mission Using a Solar Electric Spacecraft, Mar. 1, 1968 (JPL internal document).

would be the sample average life. One way of stating this approximation in a precise mathematical form based on the strong law of large numbers is as follows: Consider the sequence,  $S_1, S_2, \dots, S_n$ , where  $S_n$  is the sample average of the life of  $n$  thrusters. Each succeeding  $S_n$  can be found by testing an additional thruster. Then, with probability one, the sequence will have a limit equal to  $1/\alpha$ .

Of course, the real law governing the reliability of a thruster will show the effects of wear after an interval of time. Therefore, the actual expected life might be considerably less than  $1/\alpha$ .

## 2. Calculation of Mission Failure Probability

Let us denote by  $t_0$  the starting time for the engine and by  $t_1, t_2, t_3$ , and  $t_4$  the times when modes 1, 2, 3, and 4 are terminated. For mode 1, we require at least one of the following possible combinations of thrusters that contain a diagonal pair (Fig. 2):

1, 3, 2, 4  
1, 3, 2, 5  
1, 3, 4, 5  
2, 4, 1, 5  
2, 4, 3, 5

For mode 2, we have the following allowable combinations:

1, 3, 2  
1, 3, 4  
1, 3, 5  
2, 4, 1  
2, 4, 3  
2, 4, 5

In mode 3, we have the following allowable combinations:

1, 3  
2, 4

In mode 4, any thruster can be used since attitude control is by means of a cold gas system.

The mathematical problem is to calculate the probability that in at least one of these modes, no working combination of thrusters exists that have been listed. For instance, if thrusters 3 and 4 both fail before  $t_3$ , the end of

the third mode, then the mission is considered to be a failure. A modification of this problem is to assume that thruster 5 has been eliminated from the configuration, so that only those combinations not containing this thruster are allowed.

The probability  $p(t)$  that a thruster is operating at time  $t$  is given in terms of  $\alpha$ :

$$p(t) = e^{-\alpha t} \quad (1)$$

The probability of mission failure is then calculated as a function of  $\alpha$  and of the times  $t_i, i = 1, 2, 3, 4$ . This can be done by considering the (assumed) independent events of the failure of different thrusters.

For bookkeeping purposes, consider a fifth mode following the first four in which a thruster must fail if it does not fail during one of the first four modes. The probability of failure of a thruster at the  $i$ th mode is given by

$$\exp(-\alpha t_{i-1}) - \exp(-\alpha t_i) \quad i = 1, 2, 3, 4, 5 \quad (2)$$

where  $t_5 = \infty$ . The joint probability of the failure of thrusters in different modes is given by

$$\prod_{j=1}^n [\exp(-\alpha t_{i-1,j}) - \exp(-\alpha t_{i,j})] \quad (3)$$

where  $t_{i,j}$  is the  $i$ th mode at which the  $j$ th thruster fails, and  $n$  is the number of thrusters employed (four or five).

The method of calculation of the mission failure probability is to consider all possibilities of failure ( $5^5 = 3125$  cases for five thrusters and five modes) and to check each case to see whether it is possible to choose working thrusters for each of the first four modes. Since these events are exhaustive and mutually exclusive, the probability of mission success or failure is found by adding the probabilities of the corresponding elementary events.

The machine program for this computation has been carefully checked. For instance, the probability of mission failure and success are calculated independently, and it is verified that they add to one.

## 3. Numerical Results

Figure 3 shows the probability of mission failure for a range of  $\alpha$ . From this figure we see that  $\alpha$  should be of the order of  $10^{-4}$  for the mission to have an acceptably low probability of failure. This would correspond to a

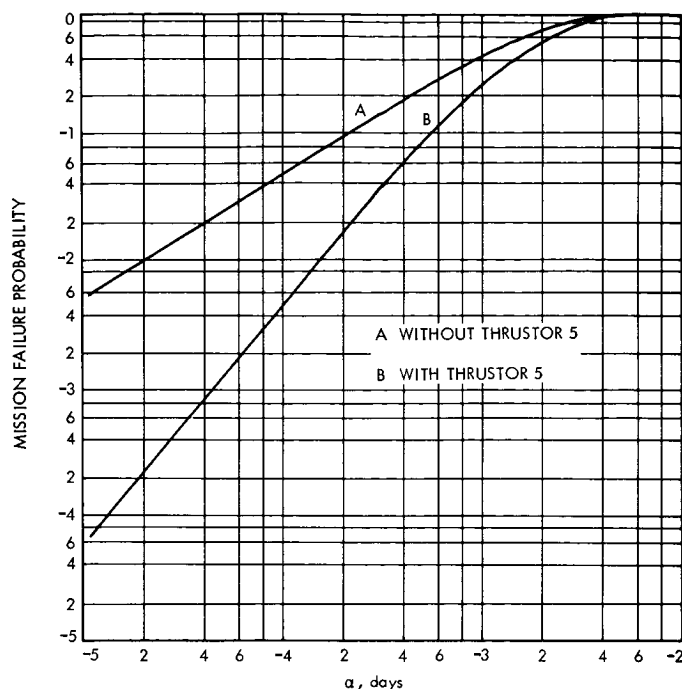


Fig. 3. Probability of mission failure

thrustor having an expected lifetime of 10,000 days if the exponential law is valid for times greater than the nominal duration of the mission. As was pointed out, the actual expected life may be considerably lower due to wear. For  $\alpha = 10^{-4}$  the probability of failure for a thrustor at the end of 470 days is approximately 0.046. For 10,000 h, or about 417 days, the probability of failure is about 0.041. Figure 4 illustrates how changes in the switching times, or times at which the modes change, influence the mission reliability.

#### D. Einstein's General Theory of Relativity: Measuring the Sun's Angular Momentum, R. W. Davies and H. Lass

##### 1. Introduction

The rotation of the sun yields an additional term to the static Schwarzschild line element of general relativity theory. The bending of light and the perihelion advance of the planet Mercury are not appreciably affected due to this additional term. The presence of this non-irrotational term makes itself known whenever light paths describe a closed orbit about the sun in the same sense that  $\oint d\phi = 2\pi$  whenever a closed path encloses the origin just once, whereas  $\oint d\phi = 0$  if the closed orbit does not surround the origin,  $r$  and  $\phi$  being polar coordinates.

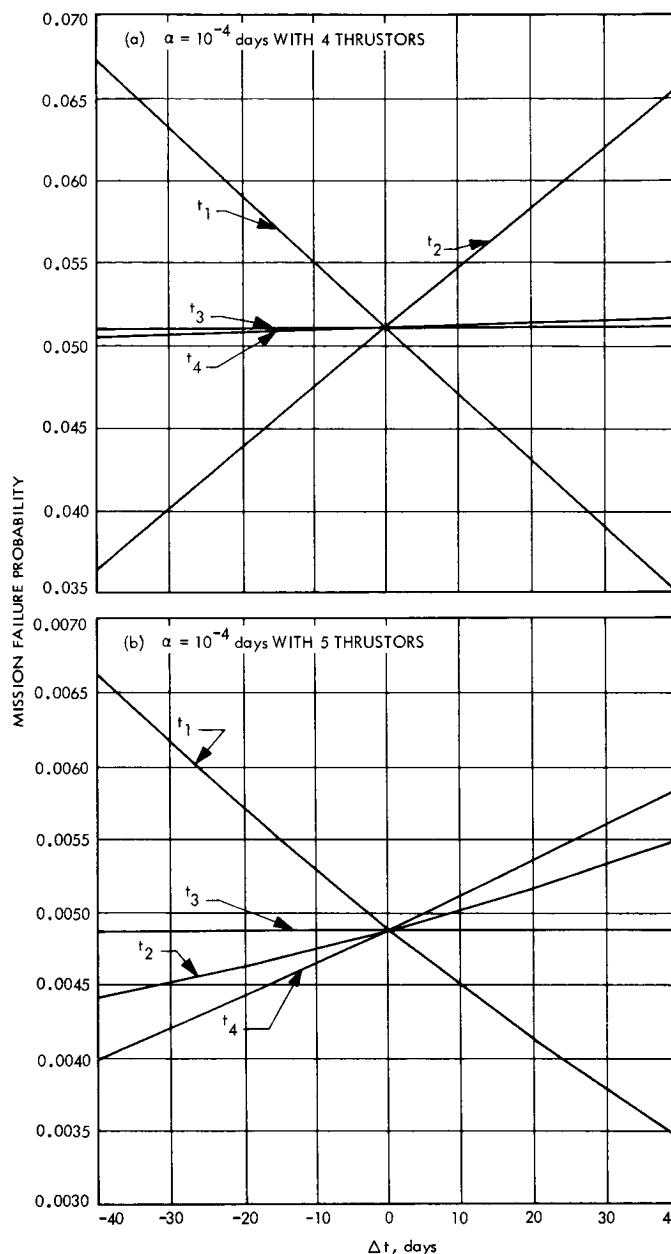


Fig. 4. Sensitivity of mission failure probability to switching times

##### 2. Discussion

The metric due to a uniform body having constant angular velocity  $\Omega$  about the  $z$ -axis is approximately

$$ds^2 = c^2 \left( 1 - \frac{2GM}{c^2 r} \right) dt^2 - \frac{dr^2}{1 - \frac{2GM}{c^2 r}} - r^2 d\phi^2 + \frac{8}{5} \frac{GMb^2 \Omega}{c^2 r} d\phi dt \quad (1)$$

in the equatorial plane of the body,  $r$  and  $\phi$  being the polar coordinates, and  $b$  the radius of the sphere with  $r \geq b$ .

The differential equations for the geodesics in the equatorial plane are obtained from

$$\left. \begin{aligned} \frac{d}{ds} \left( \frac{\partial L}{\partial \dot{\phi}} \right) &= \frac{\partial L}{\partial \phi} \\ \frac{d}{ds} \left( \frac{\partial L}{\partial \dot{r}} \right) &= \frac{\partial L}{\partial r} \\ \frac{d}{ds} \left( \frac{\partial L}{\partial \dot{t}} \right) &= \frac{\partial L}{\partial t} \end{aligned} \right\} \quad (2)$$

with

$$L = \frac{1}{2} \left[ \frac{\dot{r}^2}{1 - \frac{2GM}{c^2 r}} + r^2 \dot{\phi}^2 - \frac{8}{5} \frac{GMb^2 \Omega}{c^2 r} \dot{\phi} \dot{t} - c^2 \left( 1 - \frac{2GM}{c^2 r} \right) \dot{t}^2 \right] \quad (3)$$

such that  $\dot{r} = dr/ds$ , etc.

From  $\partial L / \partial \phi = 0$ , we obtain the first integral

$$r^2 \dot{\phi} - \frac{4}{5} \frac{GMb^2 \Omega}{c^3 r} c \dot{t} = \frac{h}{c} \quad (4)$$

Another first integral is obtained from the fact that  $\partial L / \partial \dot{t} = 0$ , so that

$$\left( 1 - \frac{2GM}{c^2 r} \right) c \dot{t} + \frac{4}{5} \frac{GMb^2 \Omega}{c^3 r} \dot{\phi} = k \quad (5)$$

Solving Eqs. (4) and (5) for  $r^2 \dot{\phi}$  and  $c \dot{t}$  yields, to a first approximation,

$$\left. \begin{aligned} r^2 \dot{\phi} &= \frac{h}{c} + \frac{4}{5} \frac{GMb^2 \Omega}{c^3 r} k \\ c \dot{t} &= \frac{k - \frac{4}{5} \frac{GMb^2 \Omega}{c^3 r^3} h}{1 - \frac{2GM}{c^2 r}} \end{aligned} \right\} \quad (6)$$

It follows that

$$\begin{aligned} \frac{dr}{ds} = \frac{dr}{d\phi} \frac{d\phi}{ds} &= \frac{1}{r^2} \frac{dr}{d\phi} \left[ \frac{h}{c} + \frac{4}{5} \frac{GMb^2 \Omega}{c^3 r} k \right] \\ &= - \frac{du}{d\phi} \left( \frac{h}{c} + \frac{4}{5} \frac{GMb^2 \Omega}{c^3} k u \right) \end{aligned} \quad (7)$$

with  $u = 1/r$ .

Dividing Eq. (1) by  $ds^2$ , and making use of Expression (6) and Eq. (7), yields, to a first approximation,

$$\begin{aligned} \frac{1 - \frac{2GM}{c^2} u}{h^2} &= \left( \alpha - \frac{4}{5} \frac{GMb^2 \Omega}{c^4} u^3 \right)^2 - \left( \frac{du}{d\phi} \right)^2 \left( \frac{1}{c} + \frac{4}{5} \frac{GMb^2 \Omega}{c^3} \alpha u \right)^2 \\ &\quad - u^2 \left( 1 - \frac{2GM}{c^2} u \right) \left( \frac{1}{c} + \frac{4}{5} \frac{GMb^2 \Omega}{c^3} \alpha u \right)^2 \\ &\quad + \frac{8}{5} \frac{GMb^2 \Omega}{c^3} u^3 \left( \frac{1}{c} + \frac{4}{5} \frac{GMb^2 \Omega}{c^3} \alpha u \right) \left( \alpha - \frac{4}{5} \frac{GMb^2 \Omega}{c^4} u^3 \right) \end{aligned} \quad (8)$$

with  $\alpha = k/h$ .

By virtue of Eq. (6), a first approximation to  $\alpha = k/h$  is given by

$$\frac{1}{\alpha} = \frac{1}{1 - \frac{2GM}{c^2 R}} R \left( R \frac{d\phi}{dt} \right)_{r=R} \quad (9)$$

with  $R$  the radial coordinate of closest approach to the sun.

Now we are interested in the motion of light for which  $ds = 0$ . This means that  $h = \infty$ ,  $k = \infty$ , with

$$\alpha = \lim_{\substack{h \rightarrow \infty \\ k \rightarrow \infty}} \frac{k}{h}$$

Returning to the metric of Eq. (1) with  $ds^2 = 0$ , and considering the closest approach of a light path to the sun such that  $r = R$ ,  $dr/dt = 0$ , we find, to a first approximation, that

$$R \left( \frac{d\phi}{dt} \right)_{r=R} = c \left( 1 - \frac{GM}{c^2 R} \right) \quad (10)$$

since

$$\frac{8}{5} \frac{GMb^2 \Omega}{cR^2} < \frac{2GM}{R} \quad (11)$$

Thus, from Eqs. (9) and (10) we find that the value of  $\alpha$  for light photons is given by

$$\alpha \approx \frac{1 - \frac{GM}{c^2 R}}{cR} \quad (12)$$

Setting  $h = \infty$  and neglecting higher order terms, Eq. (8) yields for the motion of light, the first integral

$$\left( \frac{du}{d\phi} \right)^2 + u^2 = \frac{2GM}{c^2} u^3 + u_0^2 \left( 1 - \frac{2GM}{c^2} u_0 \right) \quad (13)$$

with  $u_0 = 1/R$ , and  $du/d\phi = 0$  for  $u = u_0$  (closest approach). (Note that Eq. 13 is independent of  $\Omega$ .) However, the proper time for light to traverse a closed path around the sun will depend on whether the motion is clockwise or counter-clockwise because of the presence of the  $d\phi dt$  term in the metric of Eq. (1), setting  $ds^2 = 0$ .

From Eq. (13), we obtain

$$d\phi^2 = \frac{du^2}{u_0^2 - u^2 + \frac{2GM}{c^2} (u^3 - u_0^3)} \quad (14)$$

Furthermore,  $dr = (-1/u^2) du$ , so that Eq. (1) becomes (setting  $ds^2 = 0$ )

$$\begin{aligned} \left( 1 - \frac{2GM}{c^2 u} \right) c^2 dt^2 &= \frac{1}{1 - \frac{2GM}{c^2} u} \frac{du^2}{u^4} \\ &+ \frac{1}{u^2} \frac{du^2}{u_0^2 - u^2 + \frac{2GM}{c^2} (u^3 - u_0^3)} \\ &\mp \frac{8}{5} \frac{GMb^2 \Omega}{c^3} \frac{u_0}{u} \frac{du^2}{u_0^2 - u^2} \end{aligned} \quad (15)$$

by virtue of the fact that

$$c dt \approx \frac{u_0}{u^2} \frac{|du|}{(u_0^2 - u^2)^{1/2}}$$

The negative sign in the last term of Eq. (15) applies for the case  $d\phi > 0$  and the plus sign for the case  $d\phi < 0$ .

The additional proper time for light travel due to the rotational term in the metric can be shown to be

$$d\tau = \mp \frac{4}{5} \frac{GMb^2 \Omega}{c^4} \frac{u |du|}{(u_0^2 - u^2)^{1/2}} \quad (16)$$

Let us consider the case (unrealizable) for which the earth (E) and two satellites ( $S_1$  and  $S_2$ ) are at rest at 1 AU from the sun, the three objects forming the vertices of an equilateral triangle. The difference in time for light traveling from E to  $S_1$  to  $S_2$  and back to E as compared to light traveling from E to  $S_2$  to  $S_1$  and back to E, is

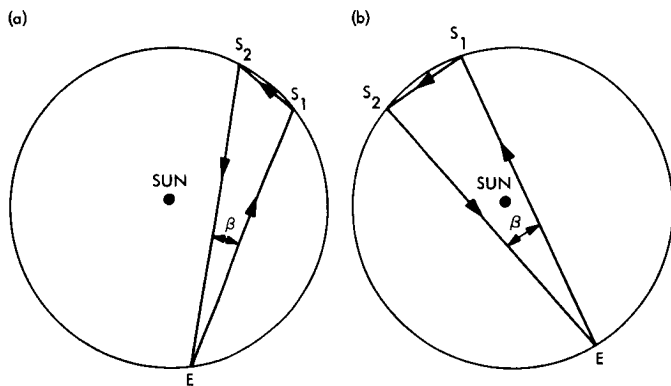
$$\begin{aligned} \tau &= \frac{48}{5} \frac{GMb^2 \Omega}{c^4} \int_{1/2 u_0}^{u_0} \frac{u du}{(u_0^2 - u^2)^{1/2}} \\ &= \frac{24}{5} 3^{1/2} \frac{GMb^2 \Omega u_0}{c^4} \end{aligned} \quad (17)$$

In cgs units, we have

$$\begin{aligned} M &\approx 2 \times 10^{33} \text{ gm} \\ b &\approx 7 \times 10^{10} \text{ cm} \\ G &\approx 6.7 \times 10^{-8} \text{ cm}^3/(\text{gm} - \text{s}^2) \\ \Omega &\approx 3 \times 10^{-6} \text{ rad/s} \\ 1 \text{ AU} &\approx 1.5 \times 10^{13} \text{ cm} \\ c &\approx 3 \times 10^{10} \text{ cm/s} \end{aligned}$$

Substituting into Eq. (17), with  $u_0 \approx 2/1 \text{ AU}$ , yields a value of  $\tau$  of the order of  $10^{-12} \text{ s}$ . If the closest approach of the light rays from the center of the sun is two sun radii, then  $\tau$  is of the order of  $10^{-10} \text{ s}$ . Unfortunately, this effect is far overshadowed when one takes into account the motion of the earth and the satellites.

If the two satellites are close to each other (Fig. 5a), the additional time effect due to the sun's rotation may



**Fig. 5. Effect of sun rotation on light travel time between earth and satellites**

be ignored since  $\phi$  increases from E to  $S_1$  to  $S_2$ , and decreases from  $S_2$  to E. When the configuration is that shown in Fig. 5b, the total effect of the sun's rotation comes into play since  $\phi$  runs through a complete cycle of  $2\pi$  radians.

### 3. Conclusion

Accurate time observations for the transit time for light to traverse the closed path  $ES_1S_2E$  conceivably could lead to a test of Einstein's general theory of relativity. Alternately, with the assumption of the validity of general relativity theory, the experiment offers a potential means to determine the total angular momentum of the solar mass.

## Subject Index

Subject	Pages	Subject	Pages
<b>Antennas and Transmission Lines</b>		<b>Electronic Components and Circuits (contd)</b>	
reflector surface materials for		400-W dc-to-ac inverter for power	
spacecraft antennas . . . . .	99-106	subsystem on Thermoelectric Outer-	
radiation characteristics of Mars rough		Planet Spacecraft (TOPS) . . . . .	130-132
lander antennas . . . . .	106-109	liquid amalgam electrodes . . . . .	133-136
measurement of fields produced by a dipole		<b>Energy Storage</b>	
in a cylindrical inhomogeneous plasma . . . . .	109-111	gravity effects on silver-zinc cells . . . . .	132-133
spacecraft antenna pointing loss:		liquid amalgam electrodes . . . . .	133-136
optimization criterion for illuminating		electrical performance of mercury-	
circular antenna apertures . . . . .	111-120	cadmium cells . . . . .	137-138
optimization of space antenna concepts . . . . .	171-177	sterilizable battery development . . . . .	213-216
<b>Apollo Project</b>		<b>Fluid Mechanics</b>	
solar plasma experiment . . . . .	18-24	gage transformation of the second	
<b>Chemistry</b>		post-Newtonian equations of	
role of CO <sub>2</sub> in Mars and Venus		hydrodynamics in general relativity . . . . .	13-15
atmospheres . . . . .	8-13	parallel-plate viscometer technique . . . . .	216-219
ion cyclotron resonance of ion-molecule		<b>Information Theory: Coding</b>	
reactions in methane-ammonia		analysis of a digital single-channel	
mixtures . . . . .	15-18	command system . . . . .	25-31
relationship between index of refraction		joint probability density of true and	
and molecular weight . . . . .	219-222	estimated phase delay in sequential	
refractive index to molecular weight		ranging system . . . . .	31-33
relationships for isobutylene telomers		behavior of phase-locked loop during	
and polymers . . . . .	222-226	temporary loss of signal . . . . .	33-37
electronic conductivity of elastomeric		choosing frequencies for squarewave	
ionenes . . . . .	226-235	subcarriers . . . . .	39-42
<b>Computer Programs</b>		signal design for fast-fading channels . . . . .	42-48
receptance coupling program . . . . .	178-179	error-correcting prefixes for establishing	
<b>Computers</b>		word synchronization . . . . .	48-50
negative radix conversion for computers		capacity of graphs . . . . .	51-54
indifferent to signs of numbers . . . . .	54-60	<b>Information Theory: Decoding</b>	
digital command system development . . . . .	126-129	sensitivity of sequential decoders to	
enhancement of real-time data processing		noise correlation . . . . .	37-39
and memory space requirements through		<b>Lunar Exploration, Advanced</b>	
forward smoothing of sequentially		experiments for manned portion of a lunar	
correlated processes . . . . .	253-257	roving vehicle mission . . . . .	1-4
<b>Control and Guidance</b>		<b>Lunar Interior</b>	
Thermoelectric Outer-Planet Spacecraft		gravimetric experiments for manned portion	
(TOPS) attitude-control subsystem		of a lunar roving vehicle mission . . . . .	1-4
simulation instrumentation . . . . .	149-151	<b>Lunar Motion</b>	
flexible structure effects on		numerical integration of the motion	
attitude-control sensors . . . . .	151-155	of the moon . . . . .	252-253
failure detection system for solenoid-		<b>Lunar Surface</b>	
actuated gas jet valves . . . . .	155-158	experiments for manned portion of a lunar	
Wiener-type filtering for multivariable		roving vehicle mission . . . . .	1-4
nonlinear processes . . . . .	158-164	<b>Masers and Lasers</b>	
<b>Electronic Components and Circuits</b>		masers for frequencies above S-band	
10.6 $\mu$ m heterodyne radiometer for		for space communications . . . . .	90-93
optical communications . . . . .	95-99		

## Subject Index (contd)

Subject	Pages	Subject	Pages
<b>Masers and Lasers (contd)</b>		<b>Mechanics</b>	
vibration and deformation measurements		flexible structure effects on	
using holography techniques . . . . .	187-190	attitude-control sensors . . . . .	151-155
<b>Materials, Metallic</b>		receptance coupling computer program . . . . .	178-179
reflector surface materials for		maximum dynamic response of structures . . . . .	179-183
spacecraft antennas . . . . .	99-106	<b>Optics</b>	
long-life spacecraft pressure vessel		10.6 $\mu$ m heterodyne radiometer for	
materials . . . . .	167-169	optical communications . . . . .	95-99
effect of space environment on spacecraft		microstrain measuring techniques for use	
thermal control materials . . . . .	169-170	on photoheliograph mirror material . . . . .	165-167
liquid propulsion system materials		<b>Particle Physics and Nuclear Reactions</b>	
compatibility . . . . .	250-251	ion cyclotron resonance of ion-molecule	
<b>Materials, Nonmetallic</b>		reactions in methane-ammonia mixtures . . . . .	15-18
relationship between index of refraction		<b>Planetary Atmospheres</b>	
and molecular weight of polymers . . . . .	219-222	gas chromatograph/mass spectrometer for	
refractive index to molecular weight		analysis of Jupiter atmosphere . . . . .	5-7
relationships for isobutylene		role of CO <sub>2</sub> in Mars and Venus atmospheres . . . . .	8-13
telomers and polymers . . . . .	222-226	<b>Planetary Exploration, Advanced</b>	
electronic conductivity of		gas chromatograph/mass spectrometer for	
elastomeric ionenes . . . . .	226-235	analysis of Jupiter atmosphere . . . . .	5-7
viscoelastic behavior of elastomers		radiation characteristics of Mars rough	
undergoing scission reactions . . . . .	235-237	lander antennas . . . . .	106-109
long term aging of elastomers: chemical		<b>Plasma Physics</b>	
stress relaxation and other studies of		measurement of fields produced by a dipole	
peroxide-cured styrene-butadiene rubber . . . . .	237-241	in a cylindrical inhomogeneous plasma . . . . .	109-111
liquid propulsion system		comparison of theoretical and experimental	
materials compatibility . . . . .	250-251	deflection of a blown arc . . . . .	242-246
<b>Mathematical Sciences</b>		<b>Power Sources</b>	
capacity of graphs . . . . .	51-54	400-W dc-to-ac inverter for power	
negative radix conversion . . . . .	54-60	subsystem on Thermoelectric	
solution of a Toeplitz set of linear equations . . . . .	65-73	Outer-Planet Spacecraft (TOPS) . . . . .	130-132
correlations between a given data		effects of space environment on solar cells . . . . .	136-137
sequence and cyclic shifts of a			142-145
pseudo-noise sequence . . . . .	73-77	electric propulsion power conditioning . . . . .	138-142
a class of sequence permutations suitable		reference design criteria for Thermoelectric	
for voice scrambling . . . . .	78-81	Outer-Planet Spacecraft (TOPS)	
spectrum analysis of scrambling		power source . . . . .	145-147
permutation . . . . .	81-89	liquid-metal magnetohydrodynamic power	
application of WKB (Wentzell-Kramers-		conversion, a candidate power source for	
Brillouin) approximation to scalar		nuclear-electric propulsion . . . . .	246-249
wave equation . . . . .	109-111	<b>Propulsion, Electric</b>	
Wiener-type filtering for multivariable		power conditioning . . . . .	138-142
nonlinear processes . . . . .	158-164	comparison of theoretical and experimental	
forward smoothing of sequentially		deflection of a blown arc . . . . .	242-246
correlated processes . . . . .	253-257	liquid-metal magnetohydrodynamic power	
calculation of probabilities of mission		conversion, a candidate power source for	
failure for solar-electric propulsion		nuclear-electric propulsion . . . . .	246-249
systems using varying numbers			
of thrusters . . . . .	257-259		

## Subject Index (contd)

Subject	Pages	Subject	Pages
<b>Propulsion, Electric (contd)</b>		<b>Structural Engineering</b>	
calculation of probabilities of mission		maximum dynamic response of structures . . . . .	179-183
failure for solar-electric propulsion		deformation modes of a simple rotating	
systems using varying numbers		structure . . . . .	184-187
of thrusters . . . . .	257-259	vibration and deformation measurements	
		using holography techniques . . . . .	187-190
<b>Propulsion, Liquid</b>		<b>Sun</b>	
long-life spacecraft pressure		<i>Apollo</i> solar plasma experiment . . . . .	18-24
vessel materials . . . . .	167-169	relationship of sun's rotation to	
propulsion system analysis for thrusters		general relativity theory . . . . .	259-262
of differing performance characteristics . . . . .	207-212		
propulsion system materials compatibility . . . . .	250-251	<b>Telemetry and Command</b>	
<b>Propulsion, Solid</b>		digital single-channel command system . . . . .	25-31
heat transfer in a burning solid propellant		S/X-band experiment for <i>Viking</i> orbiter	
with oscillating surface temperature		system . . . . .	93-95
and surface regression . . . . .	191-199	radio-frequency subsystem for	
combustion termination of solid propellant		Thermoelectric Outer-Planet	
by rapid depressurization . . . . .	200-206	Spacecraft (TOPS) . . . . .	121-125
propulsion system analysis for thrusters of		digital command system development . . . . .	126-129
differing performance characteristics . . . . .	207-212		
electronic conductivity of		<b>Temperature Control</b>	
elastomeric ionenes . . . . .	226-235	effect of space environment on spacecraft	
viscoelastic behavior of elastomers		thermal control materials . . . . .	169-170
undergoing scission reactions . . . . .	235-237		
long term aging of elastomers: chemical		<b>Test Facilities and Equipment</b>	
stress relaxation and other studies of		equipment for testing long-life spacecraft	
peroxide-cured styrene-butadiene		pressure vessel materials . . . . .	167-169
rubber . . . . .	237-241	vibration and deformation measurements	
		using holography techniques . . . . .	187-190
<b>Quality Assurance and Reliability</b>		parallel-plate viscometer . . . . .	216-219
failure detection system for solenoid-			
actuated gas jet valves . . . . .	155-158	<b>Thermodynamics</b>	
<b>Quantum Theory and Relativity</b>		heat transfer in a burning solid propellant	
gauge transformation of the second post-		with oscillating surface temperature	
Newtonian equations of hydrodynamics		and surface regression . . . . .	191-199
in general relativity . . . . .	13-15		
relationship of sun's rotation to		<b>Thermoelectric Outer-Planet Spacecraft</b>	
general relativity theory . . . . .	259-262	<b>(TOPS) Project</b>	
<b>Radio Astronomy</b>		radio-frequency subsystem . . . . .	121-125
time-of-arrival observations of		400-W dc-to-ac inverter for TOPS	
eleven pulsars . . . . .	60-65	power subsystem . . . . .	130-132
<b>Scientific Instruments</b>		reference design criteria for TOPS	
gas chromatograph/mass spectrometer for		power source . . . . .	145-147
analysis of Jupiter atmosphere . . . . .	5-7	Reference Design Document for Atomic	
<i>Apollo</i> solar wind spectrometer . . . . .	18-24	Energy Commission safety requirements . . . . .	148
<b>Solid-State Physics</b>		attitude-control subsystem simulation	
microstrain measuring techniques for use		instrumentation . . . . .	149-151
on photoheliograph mirror material . . . . .	165-167		
deformation modes of a simple		<b>Tracking</b>	
rotating structure . . . . .	184-187	joint probability density of true	
viscoelastic behavior of elastomers		and estimated phase delay in	
undergoing scission reactions . . . . .	235-237	sequential ranging system . . . . .	31-33
		behavior of phase-locked loop during	
		temporary loss of signal . . . . .	33-37

## Subject Index (contd)

Subject	Pages	Subject	Pages
<b>Viking Project</b>		<b>Voice Communications (contd)</b>	
S/X-band experiment for orbiter system .....	93-95	spectrum analysis of scrambling	
		permutation .....	81-89
<b>Voice Communications</b>		<b>Wave Propagation</b>	
permutation transformation for		measurement of fields produced by a dipole	
voice scrambling .....	77-78	in a cylindrical inhomogeneous plasma .....	109-111
sequence-permutation device for		optimization criterion for illuminating	
voice scrambling .....	78-81	circular antenna apertures .....	111-120